



T.C

AKSARAY ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI

**MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME YÖNTEMLERİ
KULLANILARAK BİST100 ENDEKSİ İLE EMTİA, DÖVİZ FİYATLARI
VE GELİŞMEKTE OLAN ÜLKELERİN BORSA ENDEKSLERİ
ARASINDAKİ İLİŞKİNİN İNCELENMESİ**

YÜKSEK LİSANS TEZİ

SERAP AKBULUT

DANIŞMAN

DOÇ. DR. KEMAL ADEM

Aksaray 2022

AKSARAY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI

**MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME YÖNTEMLERİ
KULLANILARAK BİST100 ENDEKSİ İLE EMTİA, DÖVİZ FİYATLARI
VE GELİŞMEKTE OLAN ÜLKELERİN BORSA ENDEKSLERİ
ARASINDAKİ İLİŞKİNİN İNCELENMESİ**

YÜKSEK LİSANSTEZİ

SERAP AKBULUT

DANIŞMAN

DOÇ. DR. KEMAL ADEM

Aksaray 2022

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren(....) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Serap

Soyadı : Akbulut

Bölümü : Yönetim Bilişim Sistemleri

İmza :

Teslim Tarihi : 31.03.2022

TEZİN

Türkçe Adı : Makine Öğrenmesi ve Derin Öğrenme Yöntemleri Kullanılarak BİST100 Endeksi ile Emtia, Döviz Fiyatları ve Gelişmekte Olan Ülkelerin Borsa Endeksleri Arasındaki İlişkinin İncelenmesi

İngilizce Adı : Investigation of the Relationship Between BIST100 Index and Commodities, Exchange prices and Stock Market Indexes of Developing Countries Using Machine Learning and Deep Learning Method.

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazarın Adı Soyadı: Serap Akbulut

İmza:

(Handwritten signature of Serap Akbulut)

T.C.
AKSARAY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

JÜRİ ONAY SAYFASI

Serap Akbulut tarafından hazırlanan “Makine Öğrenmesi ve Derin Öğrenme Yöntemleri Kullanılarak BİST100 Endeksi ile Emtia, Döviz Fiyatları ve Gelişmekte Olan Ülkelerin Borsa Endeksleri Arasındaki İlişkinin İncelenmesi” başlıklı tez çalışması aşağıdaki jüri tarafından oy birliği ile Aksaray Üniversitesi Yönetim Bilişim Sistemleri Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Doç. Dr. Kemal ADEM

İmza

Yönetim Bilişim Sistemleri, Aksaray Üniversitesi

.....

Üye: Dr. Öğr. Üyesi Rahmi BAKİ

Yönetim Bilişim Sistemleri, Aksaray Üniversitesi

.....

Üye: Dr. Öğr. Üyesi Uğur ERKAN

Bilgisayar Mühendisliği, Karamanoğlu Mehmetbey Üniversitesi

.....

Tez Savunma Tarihi: 31/03/2022

Sosyal Bilimler Enstitüsü Yönetim Kurulu’nuntarih ve
..... sayılı kararı ile onaylanmıştır.

Sosyal Bilimler Enstitüsü Müdürü

Doç. Dr. Sevilay USLU DİVANOĞLU

TEŞEKKÜR

Bu çalışma sırasında isteklerimi göz önünde bulundurarak değerli bilgilerini benimle paylaşan, zaman ayırıp büyük bir ilgiyle çalışma sürem boyunca samimiyetini esirgemeyen, anlayışlı tavrı ve motivasyonu ile desteğini daima hissettiren saygıdeğer danışman hocam Sayın Doç. Dr. Kemal ADEM'e teşekkürü bir borç bilirim. Ayrıca çalışma sırasında yüksek lisans tezime katkı sağlayarak, her durumda desteklerini ve güvenini hissettiğim aileme teşekkür ederim.

AKSARAY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME YÖNTEMLERİ
KULLANILARAK BİST100 ENDEKSİ İLE EMTİA, DÖVİZ FİYATLARI
VE GELİŞMEKTE OLAN ÜLKELERİN BORSA ENDEKSLERİ
ARASINDAKİ İLİŞKİNİN İNCELENMESİ

(Yüksek Lisans Tezi)

Serap Akbulut, Aksaray 2022

ÖZET

Döviz piyasaları, emtia piyasaları ve gelişmekte olan ülkelerin borsa endekslerinin Bist100 üzerindeki etkisi oldukça önemlidir. Ülke ekonomileri hem kendi hem de diğer ülkelerin ekonomilerine güçlü bir şekilde bağlıdır ve bu sebeple piyasalar etkilenmektedirler. Ekonomik açıdan piyasayı takip etmek kararları doğru verebilmek için tahmin yöntemleri kullanılmaktadır. Bu çalışmada makine öğrenmesi ve derin öğrenme yöntemleri kullanılarak Bist100 endeksi ile emtia, döviz fiyatları ve gelişmekte olan ülkelerin borsa endeksleri arasındaki ilişki tahmin algoritmaları kullanılarak incelenmiştir. Veri kümesi Ocak 2017 – Ekim 2021 tarihleri arasındaki kapanış verilerinden oluşmaktadır. Deneysel çalışmalarda objektifliğin sağlanması amacıyla k katlı çapraz geçerlilik modeli uygulanmıştır. Modellerin karşılaştırılmasında; Ortalama Mutlak Hata (MAE), Bağlı Mutlak Hata (RAE) , Ortalama Karesel Hata Karekökü (RMSE) kullanılmıştır. Makine öğrenmesinde IBk, Kstar, Random Committee ve Random Forest modelleri kullanılırken derin öğrenme yöntemi olarak LSTM modeli kullanılmıştır. Deneysel çalışmalar sonucunda, LSTM modelinin makine öğrenmesi modellerine göre daha iyi sonuç verdiği görülmüştür. LSTM modeli için test sonuçları incelendiğinde MAE değer 10.27, RMSE değeri 14.15 ve RAE değeri ise 6.06 ‘dır.

Bilim Kodu : 92432, 92431, 92429, 115305

Anahtar Kelimeler : Yapay Zeka, Derin Öğrenme, Veri Madenciliği, Finansal Tahmin.

Sayfa Adedi : 68

Danışman : Doç. Dr. Kemal ADEM

AKSARAY UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
INVESTIGATION OF THE RELATIONSHIP BETWEEN BIST100 INDEX
AND COMMODITIES, STOCK MARKET PRICES AND EXCHANGE
INDEXES OF DEVELOPING COUNTRIES USING MACHINE LEARNING
AND DEEP LEARNING METHOD

(M.A. Thesis)

Serap Akbulut, Aksaray 2022

ABSTRACT

The effect of foreign exchange markets, commodity markets and stock market indices of developing countries on Bist100 is very important. The economies of the countries are strongly dependent on the economies of both their own and other countries, and therefore the markets are affected. Forecasting methods are used to follow the market economically and to make correct decisions. In this study, using machine learning and deep learning methods, the relationship between Bist100 index and commodity, foreign exchange prices and stock market indices of developing countries was examined using estimation algorithms. The dataset consists of closing data between January 2017 and October 2021. In order to ensure objectivity in experimental studies, a k-fold cross-validation model was applied. In comparison of models; Mean Absolute Error (MAE), Relative Absolute Error (RAE), Root Mean Square Error (RMSE) were used. While IBk, Kstar, Random Committee and Random Forest models are used in machine learning, the LSTM model is used as a deep learning method. As a result of experimental studies, it has been seen that the LSTM model gives better results than machine learning models. When the test results for the LSTM model are examined, the MAE value is 10.27, the RMSE value is 14.15, and the RAE value is 6.06.

Science Code : 92432, 92431, 92429, 115305

Key Words : Artificial Intelligence, Deep Learning, Data Mining, Financial Estimate.

Page Number : 68

Supervisor : Assoc. Prof. Dr. Kemal ADEM

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU	İ
ETİK İLKELERE UYGUNLUK BEYANI.....	İİ
JÜRİ ONAY SAYFASI	İİİ
TEŞEKKÜR	İV
ÖZET	V
ABSTRACT.....	VI
TABLolar LİSTESİ.....	İX
ŞEKİLLER TABLOSU	X
SİMGELER VE KISALTMALAR LİSTESİ.....	XI
1. GİRİŞ.....	1
2. LİTERATÜR ÖZETİ.....	4
3. MATERYAL VE YÖNTEM	8
3.1. Kullanılan Veri Kümeleri ve Grafikler.....	8
3.1.1. Emtia Piyasaları.....	9
3.1.2. Döviz Piyasaları	11
3.1.3. Borsa Endeksleri.....	12
3.2. Makine Öğrenmesi.....	14
3.2.1. Sınıflandırma Yöntemi (Classification Method).....	15
3.2.2. Tahmin Yöntemi (Regression Method)	15
3.3.3. Kümeleme (Clustering)	16
3.3.4. Birliktelik Kuralı (Association Rule Mining).....	16
3.3.5. Boyut Azaltma (Dimensionality Reduction).....	17
3.3. MAKİNE ÖĞRENMESİ TÜRLERİ.....	17
3.3.1. Denetimli Makine Öğrenmesi.....	17
3.3.2. Denetimsiz Makine Öğrenmesi.....	17
3.3.3. Takviyeli Makine Öğrenmesi	18

3.4. MAKİNE ÖĞRENMESİNDE KULLANILAN ALGORİTMALAR	18
3.4.1. Yapay Sinir Ağı.....	18
3.4.2. Support Vector Machine (Destek Vektör Makinesi).....	22
3.4.3. IBk (K-nearestneighbor – K-en yakın komşuluk).....	24
3.4.4. Random Forest (Rastgele Orman)	25
3.4.5. Random Comittee (Rastgele Komite)	25
3.4.6. KStar (Kyıldız -K*)	26
3.4.7. Regression By Discretization (Ayrıklaştırma ile Regresyon)	26
3.4.8. Decision Table(Karar Ağacı)	27
3.5. DERİN ÖĞRENME.....	28
3.5.1. Uzun Kısa Süreli Bellek (LSTM)	29
4. DEĞERLENDİRME ÖLÇÜTLERİ.....	33
4.1. Çapraz Doğrulama (Cross Validation)	33
4.2. Ortalama Mutlak Hata (Mean Absolute Error- MAE).....	34
4.3. Ortalama Karesel Hata Karekökü (Root Mean Square Error- RMSE).....	34
4.4. Bağıl Mutlak Hata (Relative Absolute Error-RAE)	35
5. GELİŞTİRME ORTAMLARI	36
5.1. Çalışmada Kullanılan IDE/Programlama Dili/Kütüphane.....	36
6. BULGULAR.....	38
7. SONUÇ VE ÖNERİLER	48
8. KAYNAKÇA	51
9. TEZ DEĞERLENDİRME FORMU	56

TABLÖLAR LİSTESİ

Tablo 1. Veri Setinden Örnek Değerler	9
Tablo 2. Emtia Piyasaları Değer Aralıkları.....	9
Tablo 3. Döviz Piyasaları Değer Aralıkları (Dolar, Euro, Pound).....	11
Tablo 4. Borsa Endeksleri Aralıkları	12
Tablo 5. Çalışmada Kullanılan Programlama Dili.....	36
Tablo 6. Makine Öğrenmesi 5 ve 10 Çapraz Doğrulama ile Performans Karşılaştırması...	40
Tablo 7. Makine Öğrenmesi Performans Karşılaştırması.....	41
Tablo 10. Performans Değerleri.....	46

ŞEKİLLER TABLOSU

Şekil 1. Altın, Petrol, Gümüş, Doğal Gaz ve Pamuk Fiyat Değişimi	10
Şekil 2. Amerikan Doları, Euro ve İngiliz Sterlini Fiyat Değişimi	12
Şekil 3. Çin, Hindistan, Meksika, Rusya, Brezilya ve Türkiye Borsa Döviz Grafiği.....	13
Şekil 4. Makine Öğrenmesi Yöntemlerinin Kategorizasyonu	15
Şekil 5. Biyolojik Nöron	19
Şekil 6. LSTM (Unut Kapısı ve Çıkış Kapısı).....	31
Şekil 7. Değerlendirme Ölçütleri	33
Şekil 8. Çalışmanın Mimarisi	39
Şekil 9. IBk Tahmin Grafiği	42
Şekil 10. Random Committee Tahmin Grafiği.....	42
Şekil 11. Random Forest Tahmin Grafiği.....	43
Şekil 12. KStar Tahmin Grafiği	44
Şekil 13. Kullanılan Modeller ve Parametreler.....	45
Şekil 14. LSTM Tahmin Grafiği.....	47

SİMGELER VE KISALTMALAR LİSTESİ

T.C.	Türkiye Cumhuriyeti
β	Beta
MAE	Ortalama Mutlak Hata
RAE	Bağıl Mutlak Hata
RMSE	Ortalama Karesel Hata Karekökü
YSA	Koşullu Çekişmeli Üretici Ağlar
ADAM	Adaptif Moment Tahmini
RNN	Tekrarlayan Sinir Ağları
LSTM	Uzun-Kısa Vadeli Hafıza Ağları
SVM	Destek vektör makineleri
BİST100	Borsa İstanbul 100 Endeksi

1. GİRİŞ

Borsa kavramı genel olarak, alıcı ve satıcının ya da vekillerinin bir araya geldiği pazarlardır. Bu pazarlar da satıcılar/alıcılar, menkul kıymetler ile bu türden kıymet olarak kabul edilen değerler üzerinde alım satım işlemi gerçekleştirmek amacıyla bir araya gelirler. Alım satım işlemi yapılan pazarlar, belirli sürelerde ve yerlerde kurulan aynı zamanda belirli kurallarla organize edilen merkezi pazarları ifade etmektedir (Karşlı, 1994: s. 16). Borsa kavramı etimolojik açıdan Belçika'nın Bruges kentinde veya Anvers kentinde 1531 yılında ilk sürekli fuarı olan “*Van Der Buers*” ailesinin isminden türemiştir (Fertekligil, 2000: s. 3). Öncelikle 16. yüzyılda Paris ve Londra’da, 17. yüzyılda Berlin’de, 18. yüzyılda Viyana ve New York’ta 19. yüzyılda ise Brüksel, Roma, Milano, Madrid, Tokyo ve İstanbul borsaları ortaya çıkmıştır.

Borsalar hakkında yapılan çalışmalar hisse senedi fiyat davranışlarının önemli olduğunu, ayrıca borsaların piyasada bulunan bilgilerinin de birer ifadesi olduğunu göstermiştir. Söz konusu kavram piyasalar teorisinin bir parçası olmakla birlikte uygulamacılar açısından hisse senetlerinin geleceğinin tahmin edilmesi son derece önemli bir konu olmuştur. Uygulamacılar, yatırımlarını şekillendirmek ve ticari kazanç sağlamak amacıyla borsalara eğilim göstermişlerdir. Modern toplumlarda ekonomik marketler sosyal organizasyonlarda önemli bir yere sahiptir.

Finansal piyasalar, fon fazlası olan ve fon açığı olan ekonomik birimlerin karşı karşıya geldiği piyasalardır. Ülke ekonomileri hem kendi hem de diğer ülkelerin finans piyasalarına güçlü bir şekilde bağlıdır. Bu yüzden piyasalarda ortaya çıkan durumdan etkilenmektedirler. Bunun yanı sıra artık finansal piyasalar sadece stratejik yatırımcılar için değil tasarruflarını değerlendirmek isteyen bireyler için de cezbedici bir yatırım aracı haline gelmiştir. Para ve sermaye piyasalarındaki yatırımcıların amacı, finansal piyasaların hareketlerini inceleyerek analiz etmek ve piyasanın gidişatına göre doğru tahminler yaparak kontrol ettikleri bütçeyi

bu doğrultuda kullanmaktır. Bunun için borsalardaki hisse senetleri hareketlerinin dinamikleri doğru anlaşılmalıdır. Söz konusu dinamikler, önceki verilerin değerlendirilerek geliştirilmesini ve gelecek dönemlerde farklı modeller aracılığıyla hisse senetlerinin ne yönde hareket sergileyeceğinin öngörülmesini sağlamaktadır.

Gelişmekte olan ülkelerde devlet iç borçlanma senetleri piyasası, hisse senedi piyasası, emtia piyasası arasındaki ilişkisi ve fiyat değişimleri ile piyasalar arası yayılma etkileri konuları önemli bir yere sahiptir. Söz konusu ülkelerde döviz kurundaki değişimler, fiyatları da etkilemektedir. Yerli para cinsinden döviz fiyatı arttığında fiyatların genel seviyesi de artma eğilimine girer. Son yıllarda emtia piyasaları ve finansal piyasalar arasında bir etki aracı olarak, kurumsal finansal yatırımcılar ön plana çıkmıştır (Ordu Akkaya, 2020: s. 3). Kurumsal yatırımcılar geleneksel ekonomik yatırım araçlarını kullanmanın yanı sıra, emtia piyasasını da yatırım gerçekleştirmek amacıyla kullanmaya başlamıştır.

Teknoloji ve bilgisayar alanında, sistemin geliştirilebilmesi için farklı öğrenme yaklaşımlarının benimsenmesi, çoğunlukla geleneksel makine öğrenme yaklaşımının kullanılmasının yanı sıra son yıllarda teknoloji ve bilgisayar donanımı konusundaki hızlı gelişmeler ile güçlü bilgisayarlar yaygın şekilde kullanılmaktadır. Geliştirilen modern öğrenme yaklaşımı olan “*Derin Öğrenme*” ile gerçek dünyadaki uygulama alanlarına yeni yaklaşımlar getirilmiş farklı algoritmalar kullanılarak çoklu soyutlama yapısı ile verilerin temsillerinin öğrenilmesi kolaylaşmıştır. Derin öğrenme, bilgiye ulaşma süresini kısaltmakta ve en doğru bilgiye ulaştırmakta bir araç olarak kullanılmaktadır. Bununla birlikte derin öğrenme yöntemi finans alanında yeni bakış açıları oluşturarak, gelecekte piyasa mekanizması üzerine öngörülerin daha sağlam temellere oturmasını sağlayacaktır. Derin öğrenme verilerin birden fazla özellik seviyesinin ve temsillerinin öğrenilmesine dayanmaktadır. Üst düzey nitelikler, alt düzey niteliklerden türetilerek hiyerarşik bir temsil ortaya çıkartır. Söz konusu temsil, soyutlamanın farklı düzeylerine denk gelen birden çok temsil seviyesini öğrenmektedir. Temel anlamda derin öğrenme, verinin temsilden öğrenmeye dayanmaktadır.

Derin öğrenme yöntemleri, işlemin doğruluğunu sağlamak, işlem sürecini kısaltmak, karmaşık verileri ve bu verilerin arasındaki etkileşimleri otomatik bir şekilde algılayıp bu verilerin analizini yapabilmektedir. Derin öğrenme ile insan gücüyle tamamlanması pek mümkün olmayan görevler kolaylıkla yerine getirilebilecektir. Sosyal medya alanlarına veya arama motorlarına gönderilen yeni resimlerin içeriğini analiz edebilme ve tanıma gibi ya da

insanüstü dikkat ve olağanüstü bir performansa ihtiyaç duyulan görevlerde faydanın oldukça fazla olması derin öğrenmenin önemini ortaya çıkarmaktadır. Özellikle sağlık alanında doğru teşhislerin konulması ve güvenlik alanında risk analizlerinin doğruluğunu artırması örnek olarak verilebilir. Derin öğrenme yöntemleri bilgiye hızlı ve doğru bir şekilde ulaşmamızı sağladığı için finans alanında kullanılması faydalı olacaktır. Ayrıca bu yöntemlerin kullanımı büyük veri kümelerine sahip olan finansal varlıkların tasarlanmasını ve fiyatlandırılmasını kolaylaştıracaktır. Uygun portföyün oluşturulması ve finansal risk durumunda finansal tahmin öne çıkan problemin çözümüne kolaylık sağlayacaktır.

Bu çalışmada, döviz piyasaları (Dolar, Euro, Pound), emtia piyasaları (Petrol, Altın, Gümüş, Doğal Gaz, Pamuk) ve gelişmekte olan ülkelerin borsa endekslerini (Shanghai (Çin), Ipc (Meksika), Sensex (Hindistan), RTSI (Rusya), Bovespa (Brezilya) ve BİST100 (Türkiye) temel olarak Ocak 2017 - Ekim 2021 tarihleri arasındaki değerler ile veri seti oluşturulmuştur. Bu veri seti Weka ve Python programlama dili üzerinde makine öğrenmesi ve derin öğrenme yöntemleriyle analiz edilerek BİST100 piyasa hareketleri tahmin edilmektedir.

Gelişmekte olan ülkelerin borsa endekslerinin incelenmesi ve tahminlenmesine ilişkin algoritmalarının kullanıldığı bu çalışma bilişim teknolojilerinin gerçekleşmesine paralel olarak gelecekte gerçekleştirilebilecek olan çalışmalara referans niteliğinde olacaktır. Farklı makine öğrenme ve derin öğrenme yöntemlerinin birlikte kullanımı ayrıca farklı piyasalarda kullanılabilir olması ile beraber performansının test edilebilmesi ile yöntemlerin araştırılması ve geliştirilmesi sonucunda gelişmekte olan ülkelere tahminlere dayalı kısa ve uzun vadeli politikalara katkı sağlayacağı öngörülmektedir.

2. LİTERATÜR ÖZETİ

Borsa endeksi ve endeks hareketlerinin tahmini finansal zaman serilerinde karşılaşılan uygulamalardan bazılarıdır. Bunun sebebi tahmin yapmanın yarı-etkin piyasa sebebiyle zor olmasıdır. Endeksler, vadeli işlem sözleşmeleri ve finansal varlıkları fiyatları lineer olmayan davranış niteliklerine sahip olduğu için söz konusu varlıkları fiyat tahminlerinde sıklıkla makine öğrenme yöntemine başvurulması gerekmektedir.

Yapay sinir ağları (YSA), insan beyninin öğrenme yöntemini taklit etmesi sonucunda, beynin öğrenme, hatırlama, genelleme yöntemleriyle elde ettiği verilerden yeni veriler üretebilme gibi fonksiyonların gerçekleştirildiği yazılım türleridir (Öztürk ve Şahin, 2018: s. 27). 1943 yılında ilk olarak yapay sinir ağı modeli sinir hekimi olan Warren Mc Culloch ve matematikçi Walter Pitts tarafından “*Sinir Aktivitesinde Düşüncelere Ait Bir Mantıksal Hesap (A Logical Calculus of Ideas Immanent in Nervous Activity)*” başlıklı makaleyle ortaya çıkmıştır. Yapay sinir ağları biyolojik sinir ağlarını taklit eden sentetik sistemlerdir (Çalışkan, Yazıcıoğlu, Demirci ve Kuş, 2018: s. 48). Yapay sinir ağları uygulamaları çoğunlukla; tahmin, tasniflendirme, veri ilişkilendirme ve yorumlama ayrıca veri filtreleme işlemlerinde kullanılmaktadır. Günümüzde ise, yaygın olarak borsa endeksi tahmini, iflas tahmini veya şirket bono sınıflaması gibi ekonomik problemlerde karşımıza çıkmaktadır. Hisse senedi endeks değeri tahmini üzerine olduğu kadar, günlük endeks değişim yönü ile ilgili de çalışmalar yapılmaktadır. Yapay sinir sistemleri öğrenme yeteneği kapasitesini genişletirken, karmaşık finansal yapı sebebiyle tutarlı olmayan ve öngörülemez bir performans ortaya çıkarabilir. Bunun yanında sürekli ve büyük veri kayıtları sebebiyle gereksiz özellikler ayrıştırılarak veri boyutlarının azaltılması algoritmanın işlem süresini kısaltmaktadır ve daha genellenebilir sonuçlar ortaya çıkarmaktadır.

Cao ve arkadaşları (2005) tarafından gerçekleştirilen çalışma ile Çin pazarındaki hisse senedi fiyatlarının tahmin edilebilmesi üzerine çalışılmış; bu çalışmayla Şangay Menkul Kıymetler Borsası’nda işlem gören firmalar için yapay sinir ağlarının hisse senedi fiyat

hareketlerini tahmin etmedeki doğruluk payı ortaya koyulmuştur. Yazarlar Fama ve French'in modeli ile tek değişkenli ve çok değişkenli sinir ağı modellerinin tahmin gücü arasında analiz gerçekleştirmiştir. Söz konusu çalışmanın sonucunda, yapay sinir sistemlerinin yatırımcılara hisse senedi seçiminde tahmin potansiyelini geliştirme fırsatı sunduğunu ve basit bir tek değişkenli modelin, çok değişkenli modele oranla getirileri tahmin etmede çok daha başarılı olduğu ortaya çıkmıştır (Cao, Leggio, Schniederjans, 2005: s. 2499-2512).

Uğur (2009) çalışmasında teknik ve temel analiz değişkenlerinden olan finansal oranları, makroekonomik değişkenler ve hisse senedi geçmiş fiyat verileri ile sanayi, mali ve hizmet sektöründeki hisse senetleri fiyatlarını tahmin etmeye çalışmıştır. Tek yönlü ve iki yönlü sabit yönlü etkiler panel veri ve Arellano-Bond ile Arellano-Bover dinamik panel veri regresyon modelleri kullanılmıştır. Kullanılan modeller sonucunda sanayi sektörü hisse senedi fiyatlarını belirleyen değişkenler; piyasa değeri/defter değeri oranı, enflasyon oranı, faiz oranı, hisse başına kazanç, dönen varlık devir hızı olduğu tespit edilmiştir (Uğur, A. 2014: s.3).

Kara ve arkadaşları (2011) çalışmasında BİST100 endeksinin günlük fiyatlarının yönünü öngörmeye çalışmıştır. İki farklı model geliştirilerek yapay sinir ağları ve destek vektör makineleri (SVM) kullanılmıştır. Endeksin günlük fiyatlarının yönünü tahmin etmedeki performansları karşılaştırılmıştır. Yapay sinir ağlar modelinin ortalama performansı %75,74 destek vektör makinelerinin ise %72,52'lik doğruluk payı söz konusudur. Yapay sinir ağları modelinin önemli ölçüde daha iyi bir performans gösterdiği tespit edilmiştir. Veri setinde belirli teknik indikatörler kullanılmıştır (Kara ve arkadaşları, 2011: s.5311-5319).

Vamitha ve arkadaşları (1993)'ten itibaren, zaman değişkeninde ve zamanla değişmeyen birinci dereceden, yüksek dereceden, iki faktörlü ve ikili değişkenlerde kayıtları tahmin etmek, hisse senedi fiyatı vb. için bazı yöntemleri önermişlerdir (Vamitha ve arkadaşları 2012: s. 217-230). Çok değişkenli Markov Zinciri ile ilgili araştırmaları borsa tahmin sistemlerine yeni bir bakış açısı geliştirmiştir.

Takeuchi ve Lee (2013) temel bir momentum stratejisine ileri beslemeli bir sinir ağı uygulayarak tahmin performansını artırmayı hedeflemişlerdir. Önerilen modelde ABD (Amerika Birleşik Devletleri) hisse senedi piyasası tahminine dair performansının orijinal

momentum stratejisinden daha iyi olduđu ortaya koyulmuştur (Takeuchi ve Lee, 2013: s. 75).

Akal çalışmasında (2013) BİST100 endeksi ile Çin, Brezilya, Meksika, Hindistan, Güney Kore, Endonezya ve Rusya borsaları ile Kanada'nın GSPTSE borsası arasındaki ilişkileri inceleyerek verileri analiz etmiştir. Bu çalışma neticesinde kısa dönem nedensellikleri açısından BİST100 endeksi çoğunlukla Brezilya, Kanada, Meksika ve Çin borsalarının etkisinde kalırken Endonezya borsasının da BİST100 endeksi altında olduđu ifade edilmiştir. Diğer taraftan Türk borsası ile Hindistan, Güney Kore ve Rusya borsaları arasında nedensellik bulunmamıştır. Bu borsaların birbirinden bağımsız olduğunu söylemek mümkündür (Akal, 2013: s. 1).

Deng ve Yu, (2014) çalışmalarında, derin öğrenmeyi doğrusal olmayan gizli katmanlardan yararlanılan makine öğrenme tekniklerini denetimli veya denetimsiz öğrenme, dönüştürme, desen analizi ve sınıflandırma sınıfı olarak tanımlamıştır (Deng ve Yu, 2014).

Ding ve arkadaşları (2015) çalışmasında, hisse senetlerinin bilgi ve fiyat hareketlerine duyarlılıklarını ön plana çıkararak ve bu verilerden faydalanarak derin öğrenme ile S&P 500 endeksi fiyat hareketlerini tahmin etmeye çalışmıştır. Çalışma içerisinde haberleri ve olayları değerlendirerek uzun ve kısa dönemli olayları ayrı şekilde modellemişler ve fiyat hareketleri üzerinden yapılan tahminle karşılaştırmışlardır. Bu karşılaştırmalar sonucunda endeksi %6 oranında daha doğru tahmin edebildikleri sonucuna ulaşmışlardır (Ding, Zhang, Liu, & Duan, 2015).

Najafabadi ve arkadaşları (2015), yaptıkları çalışmalarda derin öğrenmeyi insanların son derece karmaşık problemler için karar verirken gözlemlene, analiz etme, öğrenme gibi yeteneğini taklit etmeyi amaçlayan, büyük miktarda veri kullanan bir makine öğrenmesi olduğunu tanımlanmışlardır.

Shen ve arkadaşları (2015), yaptıkları çalışmada gelişmiş bir derin inanç ağını ve eşlenik gradyan yöntemini birleştirmiş bu yöntem ile döviz kurlarını tahmin etmeyi önermiş ve bu yöntemi, geleneksel ileri beslemeli sinir ağı ile karşılaştırmıştır. Sonuç olarak derin öğrenme yaklaşımının geleneksel yöntemlerden daha iyi tahminde bulunduğunu göstermiştir (Shen, Chao ve Zhao, 2015: s. 243-253).

Patel ve arkadaşları (2015), eğilim deterministik veri ile hisse senetleri, makine öğrenmesi ve hisse fiyat endekslerini tahmin etmişlerdir. Çalışmalarında veri seti olarak CNX Nifty ve S&P Bombay Stock Exchange (BSE) Sensex fiyat endekslerinin (açılış, yüksek, düşük ve kapanış) fiyatları baz alınmıştır. Tahmin için Yapay Sinir Ağı (ANN), Destek Vektör Makineleri, Random Forest ve Naive-Bayes algoritmalarını kullanılmıştır. Karşılaştırılan dört tahmin modelinden Random Forest diğer üç modele göre daha başarılı sonuç verdiğini gözlemlemişlerdir (Patel ve arkadaşları, 2015: s. 268).

Fischer ve Krauss (2017), çalışmalarında Uzun Kısa Süreli Bellek (LSTM) ağlarını kullanarak son teknoloji tekniğini ile 1992-2015 tarihleri arasında S&P 500 hisselerinin verilerini kullanarak bir model oluşturmuşlardır. Oluşturdukları LSTM modelinin sonucunda anlamlı Random Forest, Deep Neural Net ve Lojistik Regresyon modellerine göre daha başarılı olduğu sonucuna ulaşmışlardır (Fischer ve Kraus, 2018: s.14)

Krauss ve arkadaşları (2017), S&P 500 hisse senetlerinin gelecekteki fiyatlarını öngörebilmek için derin sinir ağlarının, gradyan artırılmış ağaçların, rastgele ormanların ve bu metotların topluluklarının performansını analiz etmişlerdir (Krauss, Do ve Huck, 2017: s. 689-702).

Nahil ve Lyhyaoui (2018), yatırımcıların parasını doğru zamanda değerlendirmesine ve kısa vadede daha fazla fayda sağlayabilmesine yardımcı olabilecek vektör makinesi (SVM) ile ilgili yeni bir borsa tahmini yöntemi önermiştir. Sermaye yatırımı endüstrisi açısından zaman serisi tahmininde önemli bir araçtır. Aynı zamanda sınıflandırma üzerine kurulan makine öğrenimi tekniği, finansal verileri daha doğru tahmin edebilmektedir (Nahil ve Lyhyaoui, 2018: s. 161-169).

3. MATERYAL VE YÖNTEM

3.1. Kullanılan Veri Kümeleri ve Grafikler

Döviz piyasaları (Dolar, Euro, Pound), emtia piyasaları (Petrol, Altın, Gümüş, Doğalgaz, Pamuk) ve gelişmekte olan ülkelerin borsa endekslerini (Shanghai (Çin), Ipc (Meksika), Sensex (Hindistan), RTSI (Rusya), Bovespa (Brezilya) ve BİST100 (Türkiye) temel olarak oluşturulan veri setleri değerlendirilerek analiz yapılacak; bu verilerin Ocak 2017 – Ekim 2021 tarihleri arasındaki değerleri Python programlama dili kullanılarak derin öğrenme yöntemleriyle bağ kurularak, gelişmekte olan ülkelerin borsa endekslerine paralel olarak açıklanmaya çalışılacaktır. Söz konusu hisse senetlerine ait veriler “<https://tr.investing.com/>” üzerinden elde edilmiştir. Kullanılan veriler belirtilen tarihler arasındaki hafta sonraları ve diğer tatil günlerinde ayrıca gelişmekte olan ülkelerinde borsa tatil günleri borsanın kapalı olması nedeniyle ayrıştırılarak ele alınmıştır. Ayrıştırmalar sonucunda 14 öz nitelikli veri seti ile çalışma planlanmıştır. Çalışmada Borsa endeksleri, Emtia piyasaları ve Döviz Piyasalarının kapanış verileri ile veri setleri oluşturulmuştur.

Tablo 1. Veri Setinden Örnek Değerler

TARİH	ALTIN	PETROL	GÜMÜŞ	DOGALGAZ	PAMUK	DOLAR	EURO	POUND	SHANGHAI	IPC	SENSEX	RTSI	BOVESPA	BİST100
2017-01-03	1244.7	52.33	17.810	3.327	71.75	3.5944	3.7403	4.3991	3135.92	46123.36	26643.24	1189.32	61814	766.18
2017-01-04	1247.1	53.26	17.955	3.267	73.47	3.5713	3.7480	4.4005	3158.79	46587.74	26633.13	1176.69	61589	761.44
2017-01-05	1262.1	53.76	18.037	3.273	73.81	3.5927	3.8108	4.4618	3165.41	46719.99	26878.24	1180.31	62071	763.86
2017-01-06	1255.7	53.99	17.921	3.285	73.94	3.6445	3.8388	4.4780	3154.32	46071.57	26759.23	1172.59	61685	771.07
2017-01-09	1267.6	51.96	18.096	3.103	73.01	3.7129	3.9280	4.5158	3171.24	45553.51	26726.55	1160.27	61700	773.94
2017-01-10	1268.5	50.82	18.282	3.278	73.08	3.7885	3.9987	4.6132	3161.67	45886.27	26899.56	1175.01	62132	773.94
2017-01-11	1280.0	52.25	18.275	3.224	73.17	3.8647	4.0899	4.7196	3136.75	45933.65	27140.41	1155.50	62446	776.67
2017-01-12	1283.3	53.01	18.271	3.386	72.39	3.7597	3.9900	4.5718	3119.29	46060.98	27247.16	1175.28	63954	808.91
2017-01-13	1281.3	52.37	18.209	3.419	72.25	3.7230	3.9628	4.5340	3112.76	46182.43	27238.06	1158.19	63652	815.24
2017-01-17	1297.4	52.48	18.593	3.412	72.14	3.7610	4.0293	4.6687	3108.77	46002.56	27235.66	1156.47	63831	823.63
2017-01-18	1295.7	51.08	18.718	3.302	72.27	3.7966	4.0359	4.6543	3113.01	46360.63	27257.64	1151.65	64354	827.79

Veri setinden borsa verilerinin eksik olduğu günler ayrıştırılarak Ocak 2017 - Ekim 2021 tarihleri arasında 973 günlük fiyat bilgisi elde edilmiştir. Düzenlenen veri setinden 10 adet örnek Tablo 1’de yer almaktadır.

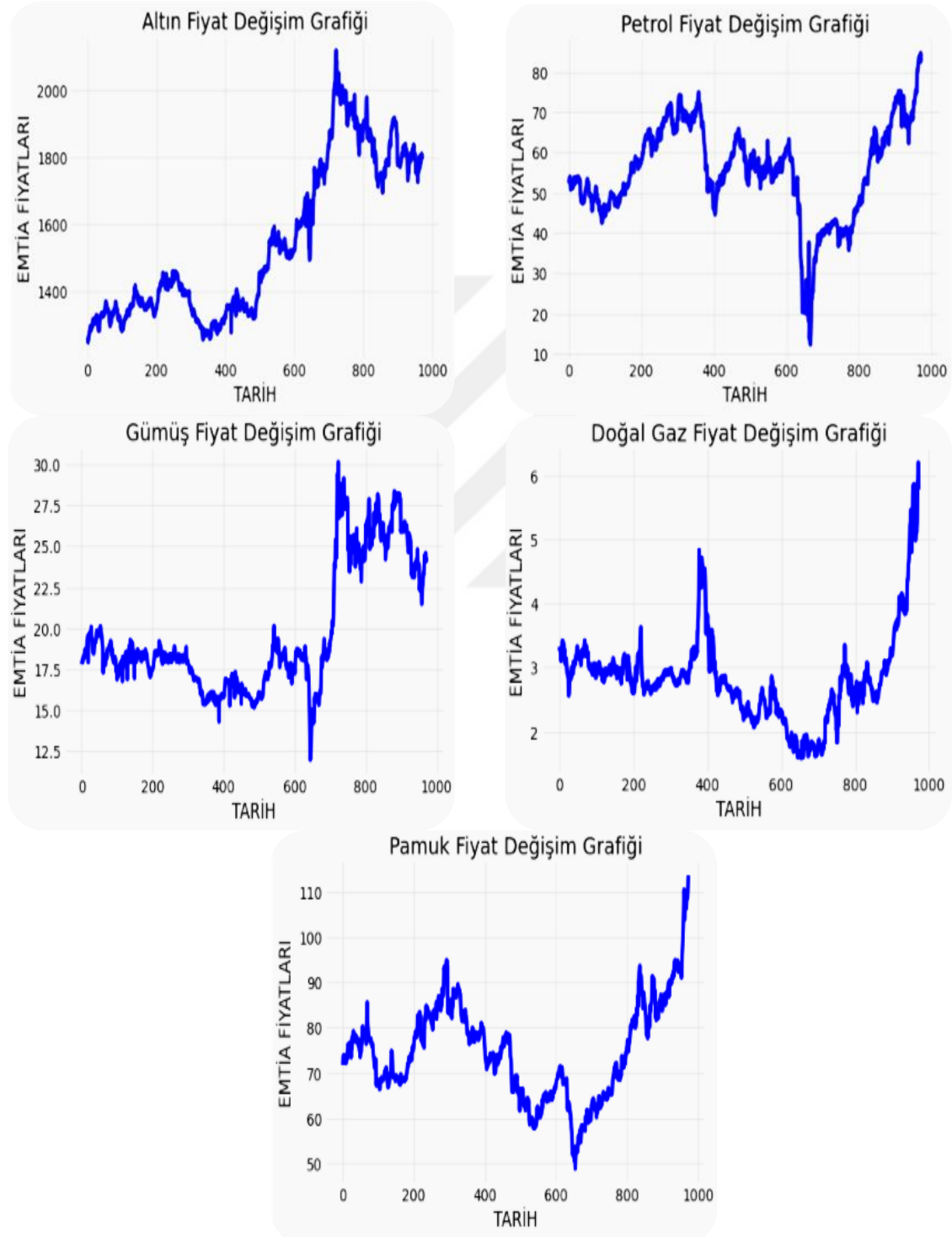
3.1.1. Emtia Piyasaları

Emtia piyasası, özellikle hammadde ya da üretilmiş durumda ticarete konu olan gümüş, altın, petrol, bakır, doğalgaz, pamuk, mısır, şeker, buğday, kahve gibi malların alınıp satılmasıyla oluşan piyasalardır. Altın Spot, Ham Petrol, Gümüş Spot, Doğalgaz ve Pamuk temel alınarak veri seti oluşturulmuştur.

Tablo 2. Emtia Piyasaları Değer Aralıkları

	count	mean	std	min	25%	50%	75%	max
ALTIN	973.0	1542.613258	230.492943	1244.7000	1344.9000	1442.7000	1772.6000	2117.1000
PETROL	973.0	55.361542	11.872654	12.3400	48.3600	56.1000	63.9100	84.6500
GÜMÜŞ	973.0	19.727333	3.890140	11.9780	17.0760	18.3100	23.3920	30.1350
DOGALGAZ	973.0	2.833024	0.706792	1.5870	2.4870	2.8010	3.0400	6.2020
PAMUK	973.0	74.705190	10.614006	48.8500	66.8600	74.1800	82.1600	113.6300

Tablo 2’de Emtia borsasındaki verilerden eşit zaman aralıklarında oluşan tanımlayıcı istatistikler incelenmiştir. İncelemeler sonucunda emtia piyasalarında yer alan Altın’ın standart sapma değeri veri setindeki diğer emtialara göre daha yüksek olduğu görülmektedir. Değişim grafikleri aşağıda yer almaktadır.



Şekil 1. Altın, Petrol, Gümüş, Doğal Gaz ve Pamuk Fiyat Değişimi

Şekil 1’de, Altın, Petrol, Gümüş, Doğal Gaz ve Pamuk Ocak 2017 – Ekim 2021 yılları arasındaki kapanışlarının fiyat değişim grafikleri gösterilmektedir. Fiyat hareketlerinin eğimleri altın ve gümüş grafiklerinde aynı yönlü hareket ederken petrol doğal gaz ve pamuk içinde kendi aralarında aynı yönde hareket ettiği görülmektedir.

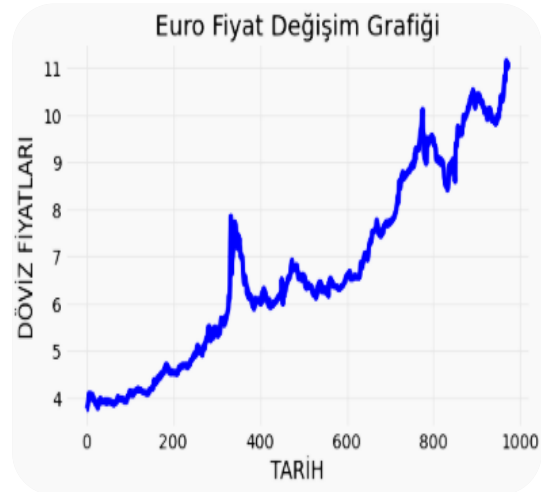
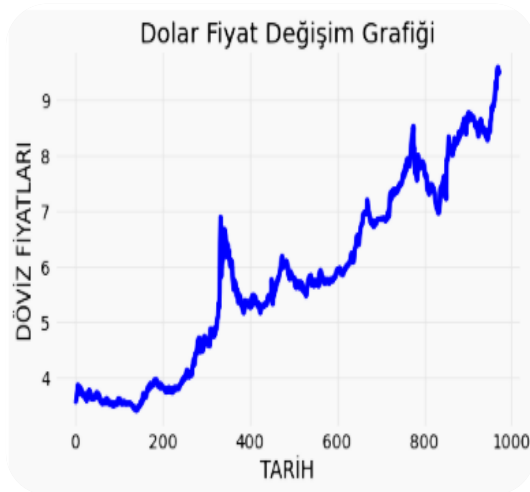
3.1.2. Döviz Piyasaları

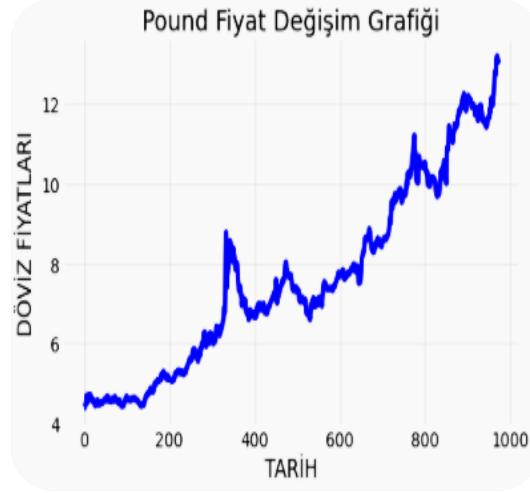
Yabancı ülke paralarının alınıp satıldığı piyasalar ya da bir ulusal para başka bir ulusal paraya çevrimi sonucu oluşan piyasalardır. Amerikan Doları / Türk Lirası, Euro / Türk Lirası, İngiliz Sterlini / Türk Lirası temel alınarak veri seti oluşturulmuştur.

Tablo 3. Döviz Piyasaları Değer Aralıkları (Dolar, Euro, Pound)

	count	mean	std	min	25%	50%	75%	max
DOLAR	973.0	5.788575	1.676877	3.4060	3.9437	5.7370	7.1570	9.5811
EURO	973.0	6.688009	2.033277	3.7403	4.8138	6.3792	8.5523	11.1528
POUND	973.0	7.614574	2.346641	4.3991	5.4603	7.3458	9.5841	13.1913

Tablo 3’te Döviz Piyasaları verilerden eşit zaman aralıklarında (Ocak 2017 ile Ekim 2021) oluşan tanımlayıcı istatistikler incelenmiştir. Standart sapmada ele alınan döviz piyasaları verilerinde en yüksek değere sahip olan döviz birimi pound olarak yer almaktadır. Değişim grafikleri aşağıda yer almaktadır.





Şekil 2. Amerikan Doları, Euro ve İngiliz Sterlini Fiyat Değişimi

Şekil 2’de, Amerikan Doları, Euro ve İngiliz Sterlini Ocak 2017 – Ekim 2021 yılları arasındaki kapanışlarının fiyat değişim grafikleri yer almaktadır. Fiyat hareketlerinin eğimleri seçilen paritelerde birbirleri ile aynı yönde hareket ettiği görülmektedir.

3.1.3. Borsa Endeksleri

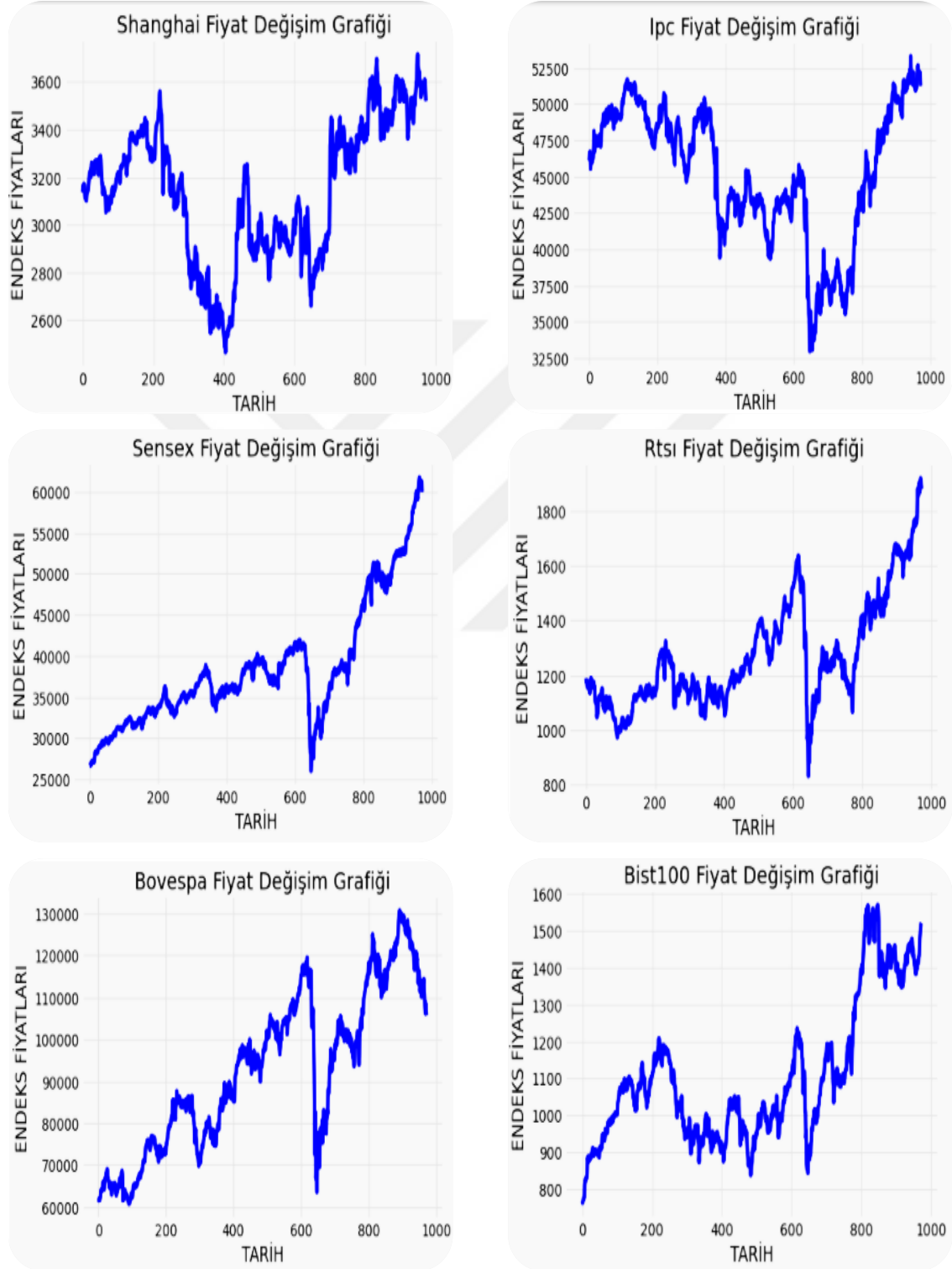
Endeks, birçok veriyi içerebilen ve bu verilerden hesaplamalar sonucunda yaklaşık değerler ile hızlı bilgiye ulaşmamızı sağlayan göstergelerdir. Bist100, Çin, Hindistan, Meksika, Rusya, Brezilya borsa endeksleri temel alınarak veri seti oluşturulmuştur.

Tablo 4. Borsa Endeksleri Aralıkları

	count	mean	std	min	25%	50%	75%	max
SHANGHAI	973.0	3141.625663	288.540981	2464.3600	2914.4800	3169.5700	3373.5800	3715.3700
IPC	973.0	45315.120144	4590.348619	32964.2200	42731.4800	45959.0400	49092.3100	53304.7400
SENSEX	973.0	38532.456804	7556.607627	25981.2400	33370.7600	36718.6000	40676.6300	61765.5900
RTSI	973.0	1271.787133	196.942396	832.2600	1128.2200	1219.8900	1387.8700	1919.5800
BOVESPA	973.0	92740.032888	18894.082714	60762.0000	76358.0000	94491.0000	107224.0000	130776.0000
BİST100	973.0	1104.052230	188.485963	761.4400	961.5800	1057.1800	1183.9900	1570.4200

Tablo 4’te veri setindeki değerlerin eşit zaman aralıklarında (Ocak 2017 ile Ekim 2021) değerleri hesaplanmıştır. Veriler incelendiğinde Rusya borsa endeksinin standart sapma

oranının diğer borsa endekslerine göre daha yüksek olduğu görülmektedir. Değişim grafikleri aşağıda yer almaktadır.



Şekil 3. Çin, Hindistan, Meksika, Rusya, Brezilya ve Türkiye Borsa Döviz Grafiği

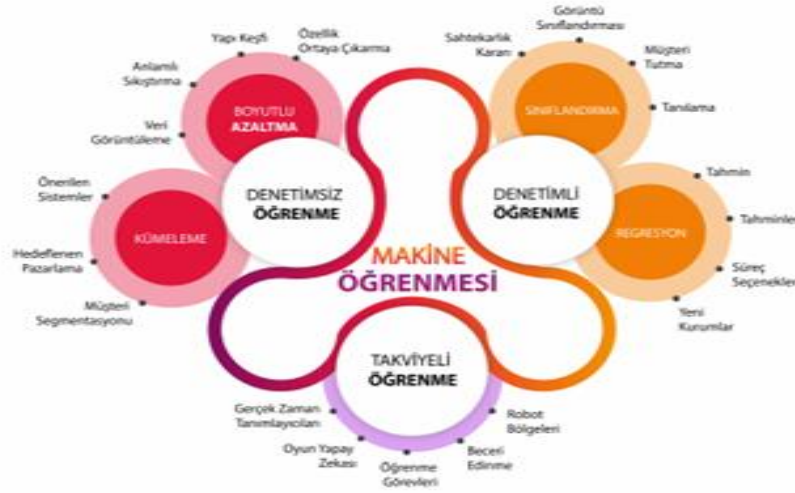
Makine öğrenmesine girdi olarak verilen Shanghai (Çin), Ipc (Meksika), Sensex (Hindistan), RTSI (Rusya), Bovespa (Brezilya) ve BIST (Türkiye)' nin Borsalarının verileri Ocak 2017 ile Ekim 2021 tarihleri arasındaki endekslerin kapanış değerlerini şekil 3' te grafik üzerinde gösterilmektedir.

3.2. Makine Öğrenmesi

Bilgi teknolojileri bilgisayarların hayatımıza girmesiyle önem kazanmaya başlamıştır. Artık nerede ise tüm bilgiler bilgisayarlar aracılığı ile veri tabanları kullanılarak kayıt altına alınmakta birçok işlem online olarak yapılabilmektedir. Bu sayede işlerin daha verimli ve sistematik ilerlemesi sağlanabilmektedir. Kayıt altına alınan verileri takip etmek kolaylaşmış, doğru bilgi ışığında gelişim hızlanmıştır. Kayıt altına alınan veriler üzerinden analizler yapılarak birçok alanda bilgiye ulaşmak daha kolay hale gelmiştir. Kamu kurumları, özel kurumlar, ülkeler veriyi analiz ederek kendilerine fayda sağlamayı amaçlamaktadır. Kayıt altına alınan veriler zamanla yüksek boyutlu veriler haline gelmiştir. Kararların doğru verilmesinde bilgi birikiminin önemi oldukça etkilidir. Veriden sağlıklı bilgiye ulaşmak için verinin toplanması, ayrıştırılması, gürültülerinin giderilmesi gerekmektedir. Bunun için makine öğreniminden faydalanılmıştır.

Makine öğrenmesi bilgisayarların insanlar gibi öğrenme yeteneğinin olabilmesi için çeşitli algoritma, modelleme ve teknikleri geliştirmek amaçlı kullanılan bilimsel bir çalışma alanıdır (Alpaydın, 2010). Arthur Samuel 1959 yılında makine öğrenmesini “Açıkça programlanmadığı halde makineler öğrenme yeteneği kazandıran disiplin” olarak tanımlamıştır. Makine öğrenmesi 1990’larda gelişmeye başlamıştır (Akpınar, 2019:s.5).

Makine öğrenimi, veriden öğrenebilen öğrendikleriyle tahminde bulunabilen algoritmalarıdır. Bilgisayarlara veriler yüklenip bu verilerden çıkarım yapan bu çıkarımla tahminde bulunabilen, programlanan bilgilerin dışındaki bilgileri de ortaya çıkarabilen ve kendi kendini geliştirebilen sistemlerdir. Makine öğrenimi, yapay zekanın bir alt alanı/alt tipidir. Bilgisayarların analitik veri modelleri oluşturmaya ve net bir şekilde kodlanmadan otomatik olarak gizli iç görüler bulmasına izin verir. Amaç, veriler üzerine matematiksel ve istatistiksel işlemler uygulayarak tahminlerden çıkarımlar yaparak öğrenme metotları ve bu metotların performansı ile ilgilenir (Alpaydın, 2010). Çözümeye ulaşmak için farklı matematiksel ve istatistiksel yöntemler kullanılır. Her bir yöntem makine öğrenmesinde kullanılan veri kümesine uygun olarak seçilmelidir.



Şekil 4. Makine Öğrenmesi Yöntemlerinin Kategorizasyonu

3.2.1. Sınıflandırma Yöntemi (Classification Method)

Eğitim veri seti üzerinden tahmin edilecek alan eğer kategorik bir değişken ise sınıflama yapılmaktadır (Köktürk ve ark.2009: s.25). Veri sınıflandırma algoritması, bir dizi gözleme en iyi uyan bir dizi alternatif arasından bir hipotez seçme işlemidir. Kadın ve erkek, hasta ve sağlıklı bireyler buna örnektir. Sınıflandırma algoritmalarına örnek olarak Naive Bayes, K-Nearest Neighbours (K-En Yakın Komşu), Decision Tree (Karar Ağacı), Random Forest (Rastgele Orman), Support Vector Machine (Destekçi Vektör Makinesi), Yapay Sinir Ağları (Artificial Neural Networks) ve Lojistik Regrasyon (Logistic Regression) verilebilir.

3.2.2. Tahmin Yöntemi (Regression Method)

Süreklilik arz eden değerleri tahmin etmek için regresyon algoritmaları kullanılır. İki ya da daha fazla nicel değişken arasındaki ilişkinin analiz sonucunu öngörmektir (Büyüköztürk,2012: s.91). Girdiler bağımsız değişken sonuç bağımlı değişkendir. Bağımsız değişkene ait her bir farklı parametre için bağımlı değişkenin tahmin edilmesi regresyonun bir amacıdır. Değişkenler arasındaki sebep-sonuç ilişkisini bulmamızı sağlar. Buna örnek olarak belirli bir bölgenin yağış miktarı ya da bir insan boyu verilebilir. Regresyon algoritmaları; Linear Regression (Doğrusal Regresyon), Logistic Regression (Lojistik Regresyon), Multiple Linear Regression (Çoklu Doğrusal Regresyon), Polynomial Regression (Polinom Regresyon)'dır.

Lineer Regresyon, Bir veya daha fazla bağımsız değişken ile bağımlı değişken ile bağımlı değişken arasındaki doğrusal ilişki aranır. Bağımsız değişken ve bağımlı değişken

arasındaki tahmin ilişkisini bulmak için kullanılır (Gültepe, 2019:s.12). Lojistik regresyon sınıflandırma problemlerinde bir veya çok ön göstergeli değişkene bağlı kategorik bağımlı değişkenlerin sonuçlarını tahmin etmede kullanılır (Ballı ve Sağbaş, 2017: s.982). Bağımlı değişken iki kategori ile ifade ediliyorsa, model “Basit Lojistik Regresyon Modeli”, ikiden daha fazla kategori ile ifade ediliyorsa, “Çoklu Lojistik Regresyon Modeli” olarak ifade edilmektedir. (Leech, Barrett ve Morgan, 2004). Örneğin bir kişinin kanser olup olmadığı, bir kişiye kredi verilip verilmeyeceği veya bir mailin spam olup olmadığı çoklu lojistik regresyon modeline örnektir.

Basit lojistik regresyon modelinde bağımlı değişkenin gözlenen değeri iki olayın (Can, Özdil, & Yılmaz, 2018) Multinomial (Çok terimli) Lojistik regresyon modeli ikiden fazla ayrı sonuca izin veren lojistik regresyonu genelleştiren bir modeldir. Yani bağımsız değişkenleri verilen, kategorik dağıtılan bağımlı değişkenlerin farklı olası sonuçlarını tahmin etmek için kullanılmaktadır (Ballı ve Sağbaş,2017:s.982).Tahmin, geçmiş ve şimdiki verilere dayanarak gelecekle ilgili tahminler yapma sürecidir ve genellikle eğilimleri analiz etmek için kullanılır.

3.3.3. Kümeleme (Clustering)

Kümeleme işlemi bir veri setinde benzer özellikleri olan verileri homojen gruplara veya kümelere ayırmaktır. Kümeleme işlemi yönteminde farklı bir kümedeki örnekleri ayrıştırarak birbirine benzer örnekleri aynı kümede toplamak amaçlanır (Günay Atbaş,2008:s.19). Kümeleme işlemi, istatistik ve makine öğrenmesi, web madenciliği, biyoloji gibi pek çok alanda kullanılmaktadır (Han, Kamber 2000: 25). Kümeleme algoritmaları; K-Means Algoritma, Gürültülü uygulamaların Yoğunluğa Dayalı Konumsal Kümelenmesi (DBSCAN), Mean Shift, Hierarchical.

3.3.4. Birliktelik Kuralı (Association Rule Mining)

Birliktelik Kuralları, 1993 yılında Agrawal, Imielinski ve Swami tarafından önerilmiştir (Ateş ve Karabatak, 2017: s. 57). Veri madenciliğinde kullanılan ilk tekniklerden birisidir. Birliktelik Kuralları, geçmiş veriler analiz edilerek büyük veri kümeleri arasındaki ilişkileri çıkaran ve nesnelerin içindeki birliktelik davranışlarının tespitini analiz eder. Herhangi bir toplamada farklı bileşenlerin bir arada olma ihtimali üzerine çalışmalar yapılır. Market sepetlerinde sıklıkla kullanılır. Örneğin bir müşteri ekmek aldıysa %80 ihtimalle yumurta da alacaktır. Birliktelik Kuralları; pazarlama eğitim, sağlık, ekonomi, telekomünikasyon, e-

ticaret, bilişim gibi pek çok alanda yaygın olarak kullanılmaktadır (Pekel,2018:s.20). Birliktelik Kuralı algoritmaları; Apriorial gorithm, Eclatal gorithm, FP-growthalgorithm.

3.3.5. Boyut Azaltma (Dimensionality Reduction)

Bir veri kümesinde kullanılan değişken veya öznitelik sayısına boyutsallık denilmektedir. Boyut azaltma, öznitelik çıkarımı ve özellik seçimi yöntemleri kullanılarak gerçekleştirilebilir. Boyut azaltma sırasında, büyük veri kümeleri içerisinde ilgisiz verileri silinerek değişken ve öznitelik sayısı azaltılır ve yöntem olarak özgün veriye ulaşmak istenmektedir. Bu işlem tüm mikro dizi veri setleri için genellikle kullanılmaktadır (Kılıçarslan, Adem ve Cömert, 2019:s.711).

3.3. Makine Öğrenmesi Türleri

3.3.1. Denetimli Makine Öğrenmesi

Denetimli öğrenmede veri etiketlenir, sınıflandırılır, girdi ve çıktı yani bağımlı ve bağımsız değişken birlikte kullanılır. Giriş verilerine eğitim verisi denir. Algoritma etiketli eğitim verisinden öğrenir. Öncelikle belirli bir miktardaki veri eğitim ve test verisi olarak bölünür. Veriye ait problem ve sonuçlar daha önceden bilinir. Eğitim verisi ile problemin çözümüne yönelik bilgiler arasında bağlantı kurularak bir algoritma oluşturulur. Sistemin eğitilmesi ve eğitilen sistemin öğrendiklerinden yola çıkarak problemin çözümüne giden yolu tahminde bulunması sağlanır. Sistemin öğrenmiş olduğu yol ile test verisini tahmin etmesi sağlanır. Girdi değişkenlerinin çıktı değişkenlerine eşleme işlevi sağlanarak yapılır.

Denetimli öğrenme; eğitim veri setinin ve test veri setinin sisteme yüklenmesi, veri setindeki her bir veri için ihtiyaç duyulan etiketlenmenin yapılması ve bu şekilde girdi veri seti ile çıktı veri seti arasında bağ kurulması ile gerçekleşir. Amaç, sonuçları bilinen veri setinden gerçekleştirilen sınıflandırma sonucunda, sonuçları bilinmeyen veri setine ilişkin etkili tahminler gerçekleştirmektir (Aydın ve Özkul, 2015: s.38).

3.3.2. Denetimsiz Makine Öğrenmesi

Denetimsiz öğrenmede veri kümesi girdi (bağımsız değişken) değişkenlerine sahiptir ancak çıktı (bağımlı değişken) değişkenleri bulunmamaktadır. Veri kümesi etiketlenmemiş olup problem ve problemin sonuçları daha önceden bilinmemektedir. Giriş verilerinde bulunan yapılar ayıklanarak bir model hazırlanır. Ayıklanan verilerden genel kurallar matematiksel bir

süreç yoluyla veya verilerin benzerliklerine göre oluşturulur. Böylece bilinmeyen bir yapı benzer özelliklerine göre bir araya getirilerek (kümelenerek) yapay öğrenme algoritmaları ile tahmin edilir.

Denetimsiz öğrenmede sisteme herhangi bir kullanıcı müdahalesi bulunmamaktadır. Sadece girdi verileri sisteme girilir fakat herhangi bir işaretleme yapılmaz. Sistem otomatik keşifler gerçekleştirir ve ilişki ağını ortaya koymaya çalışır (Alpaydın, 2010: s.11).

3.3.3. Takviyeli Makine Öğrenmesi

Takviyeli makine öğrenmede ödül en üst seviyeye çıkarılarak davranışlar öğrenilir, aracının mevcut durumuna ilişkin sıradaki en etkili işlemi belirlemesi sağlanır. Takviyeli öğrenmede yazılımlar, genellikle deneme yanılma yoluyla en iyi sonuçlara ulaşırlar ve genellikle robotikte kullanılırlar.

Etiketlenmiş az sayıda veri karşısında çok daha fazla sayıda etiketlenmemiş veri olduğu durumlarda denetimli veya denetimsiz öğrenme yetersiz kalabilir. Böyle bir durumda en uygun yöntem az sayıda etiketlenen veriden yola çıkarak etiketlenmemiş veriler hakkında bilgi sahibi olmaya çalışmaktır ve onları sınıflandırmaktır. Bu yöntem yarı denetimli öğrenme veya takviyeli makine öğrenmesi olarak nitelendirilmektedir. Denetimli öğrenmeden farkı eldeki etiketlenen veri kümesidir. Denetimli öğrenmede etiketlenmiş veri sayısı fazla, tahmin öngörüsünde bulunulmak istenen veri sayısı azken takviyeli öğrenme de tam tersi bir durum söz konusudur (Kızılkaya ve Oğuzlar,2018:s.92).

3.4. Makine Öğrenmesinde Kullanılan Algoritmalar

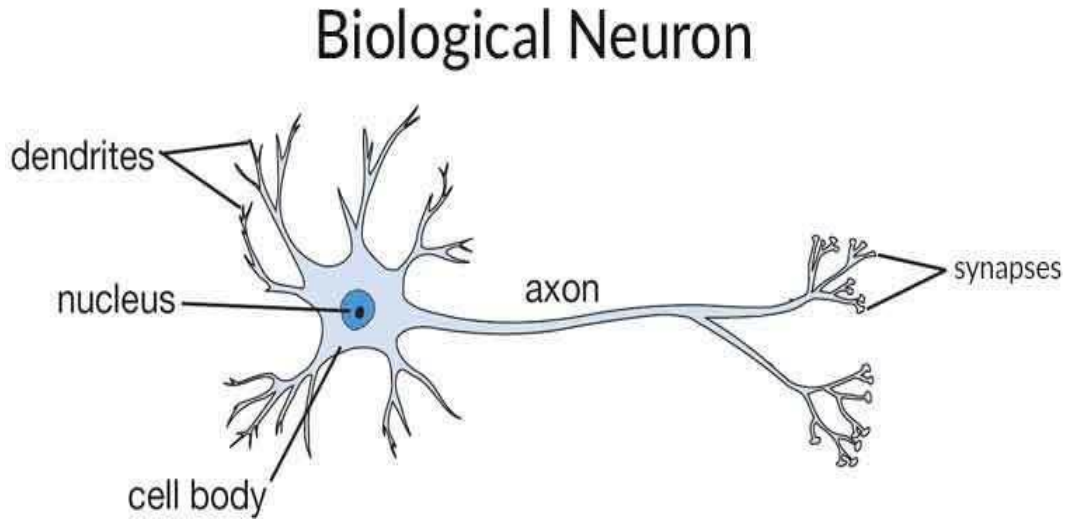
Makine öğrenmesinde kullanılan modeller; yapay sinir ağı, destek vektör makinası, IBk (K-en yakın komşuluk), Random Forest (Rastgele Orman), Random Committee (Rastgele Komite), KStar (Kyıldız), Regression By Discretization (Ayrıklaştırma ile Regresyon), Decision Table (Karar Ağacı)'dir.

3.4.1. Yapay Sinir Ağı

Yapay sinir ağları biyolojik sinir ağı yapıları örnek alınarak geliştirilmiştir. Canlıların davranışlarını inceleyip, matematiksel olarak modelleyip, benzer yapay modellerin üretilmesi ile makine öğrenmesin temelleri oluşturulmuştur. Aynı insan biyolojik yapısında olduğu gibi yapay sinir ağları vasıtasıyla makinelerin eğitilmesi, öğrenmesi ve öğrendikleriyle karar

vermesi amaçlanmaktadır (Demirci,2019:s.7). Biyolojik sinir ağı dendrit, akson, sinaps çekirdek ve miyelin tabaka bölümlerinden oluşmaktadır.

- Akson (Axon): Sistem çıkışı olan akson çıkış darbelerinin üretildiği elektriksel aktif gövdedir ve gövde üzerinde iletim tek yönlüdür.
- Dendritler (Dendrites): Diğer hücrelerden gelen işaretleri toplayan elektriksel anlamda pasif kollarıdır. Sistem girişidir(Gül,2018:s.21).
- Sinaps (Synapse): Hücrelerin aksonlarının diğer dendritlerle olan bağlantısını sağlar.
- Miyelin Tabaka (Myelin Sheath): Yayılma hızına etki eden yalıtım malzemesinden oluşmaktadır.
- Çekirdek (Nucleus): İşaretlerin periyodik olarak yeniden üretilmesini akson boyunca sağlar.



Şekil 5. Biyolojik Nöron

Yapay sinir ağları biyolojik sinir ağlarını temel alarak oluşturulmuştur. Eğitilebilir, adapte edilebilir, kendi kendine organize olup öğrenebilen ve değerlendirme yapabilen yapay sinir ağları ile insan beyninin öğrenme yapısı modellenmeye çalışılmaktadır (Demirci,2019:s.7). Bu süreçte yapay sinir ağlarında bilgiler nöronlar üzerinden işleme alınıp oluşan sinyaller nöronlar arasındaki bağlantılar ile birbirine aktarılır. Bu bağlantılar sinyalin değerlerini belirlemektedir. Nöronların her biri sonuç sağlayabilmek için girdiye bir aktivasyon fonksiyonu uygular. Bu bileşenlere göre bir yapay sinir ağı nöronlar arası bağlantı yapısı, bağlantıların ağırlıklarının belirlenmesi ve aktivasyon fonksiyonu ile ortaya çıkmaktadır (Fausett, 1994).

Yapay sinir ağıları, öğrenme yolunu taklit ederek beynin öğrenme, hatırlama ve genelleme yoluyla topladığı verilerden yeni veriler meydana getirme gibi temel fonksiyonların gerçekleştirildiği modellemelerdir. Yapay sinir ağıları lineer olmayan karmaşık fonksiyonları modellemek için beynin düşünme sürecini taklit eden sistemlerin oluşturulmasında kullanılmaktadır. Yapay sinir ağıları, tahmin, sınıflandırma, kümeleme, veri yorumlama ve filtreleme gibi farklı problemler için etkin bir araç olarak değerlendirilmiştir. Ayrıca BIST sürdürülebilirlik endeksinin tahmininde Yapay Sinir Ağları kullanılmaktadır. Yapay sinir ağıları, sinir arasındaki bağlantıların yönlerine göre İleri beslemeli (Feedforward) ve geri beslemeli (Feedback veya Recurrent) ağılar olmak üzere ikiye ayrılır (Diler, 2003: s. 68)

İleri beslemeli ağılarda, ağ içerisindeki yapay sinir hücreleri katmanlar halinde yerleştirildiğinde, ilişik katmandaki giriş verilen bilgi ağ içerisinde devamlı ileriye doğru hareket eder ve çıkış katmanından ağı çıktısı olarak belirir. İleri beslemeli ağılarda her sinir hücresi kendinden sonraki katmana bağlanabilir. Geri beslemeli ağılarda ise en az bir yapay sinir hücresinin kendi katmanındaki veya kendinden önceki katmana dönüş bağlantısı bulunmaktadır (Elmas, 2011: s. 41)

Yapay sinir ağıları geleneksel anlamda ön bilgi olmadan çıkarım gerçekleştiremez ancak yinelenen sinir ağıları belirli bir dizi çerçevesinde ve birtakım döngülerle bilginin sürekliliğini sağlamaktadır. Yinelenen sinir ağıları önceki ağ ve gizlenmiş durumu girdi olarak kullanabilir. Tahmin edilmek istenen bir durumla ilgili olarak modele verilen girdiler, modele bağımlı olarak farklı aşamalarda tekrar kullanılabilir. Yinelenen sinir ağıları girdi katmanı, içerik katmanı, gizli katmanlar ve çıkış katmanından meydana gelmektedir.

Bir problemin çözülmesi öncelikle yapay sinir ağıının topolojisine bağlıdır. Hangi sorun için hangi topolojinin uygun olduğunu belirlemek oldukça önemlidir. Probleme yönelik yapay sinir ağı topolojisinin seçimi, ağda kullanılması düşünülen öğrenme algoritmasına bağlıdır. Birçok yapay sinir ağı tek bir öğrenme algoritması ile kullanılabilirdiği için, söz konusu süreçte kullanılacak yapay sinir ağı topolojisi mecburi olarak seçilmiş olacaktır.

Ağ yapısında kullanılan katman sayısı ve katmanlarda bulunan nöron sayısı yapay sinir ağıının karmaşık yapısını belirleyen bir unsurdur. Katman sayısı ve katmanlardaki nöronların sayısı çoğaldıkça, yapay sinir ağıının işlem ve öğrenme kapasitesi ile yakınsama

süresi artmaktadır. Buna karşın ağın genelleme kabiliyeti azalarak, ağın ezberlemesine sebep olmaktadır. İhtiyacı olandan az kullanılması ise, verilerdeki desenin yeteri kadar öğrenilememesi sonucunu doğurur. Birçok problemde iki veya üç katmanlı bir ağ tatmin seviyesi yüksek sonuçlar ortaya çıkarmaktadır. İşlemci eleman sayısı ve katman sayısının belirlenmesinin en iyi yöntemi, deneme yaparak ağın performansına göre en uygun katman sayısına karar vermektir. Bu ise başlangıçtaki nöron sayısını istenilen performansa ulaşılan kadar artırarak veya istenilen performansın altına inmeden sağlanabilir.

Yapay sinir ağının tasarımında işlemci elemanlarının karakteristik özelliklerinin belirlenmesi de önemli bir konudur. Toplama fonksiyonunun ile aktivasyon fonksiyonunun seçimi, verilerin niteliklerine ve ağdan öğrenilmesi istenilen verinin tür ve yapısına bağlıdır. Aktivasyon fonksiyonları içinde en çok kullanılanlar; doğrusal, işaret, eşik, sigmoid, hiperbolik, tanjant ve lojistik fonksiyonlarıdır. Doğrusal olmayan problemlerde doğrusal olmayan aktivasyonlardan faydalanılmaktadır.

Verilerin ağa verilmeden önce normalizasyon işlemine tabi tutulması gerekmektedir. Verideki aşırı salınımların engellenmesi ve sistem performansını artırmak için bu yöntem başvurulmaktadır. Bu çerçevede logaritmik fonksiyonlar kullanılmaktadır ve genellikle verilerin $[0, 1]$ veya $[-1, 1]$ aralıklarından birine ölçeklendirilmesi gerekmektedir. Veri kalitesi aşırı problem içeren yapay sinir ağları modellerini olumsuz etkileyebilmektedir (Saraç, 2004:s.35).

İstenilen çıktı değerleri ile ağın ürettiği değerler arasındaki farkların kümülatif değerlerini performans fonksiyonları hesaplamaktadır. Hesaplanan değerler ağın, eğitim veri setinin gösterdiği desene yaklaşma seviyesini gözlemleme olanağı tanımaktadır. Bağlantının ağırlık değerleri bu bilgiler ile değiştirilmektedir. Bu sebeple performans fonksiyonları öğrenme performansını etkilemektedir. İleri beslemeli ağlarda kullanılan tipik değerlendirme ölçütü *Ortalama Karesel Hata*'dır (Gürsoy,2021:s.54). Ortalama Hata, Ortalama Karesel Hata Karekökü ve Ortalama Mutlak Hata kullanılan diğer değerlendirme ölçütlerinden bazılarıdır.

Yinelenen sinir ağları birimler arasındaki bağlantıların yönlendirilmiş bir döngü meydana getirdiği bir sınıftır. Söz konusu döngü, dinamik zamansal davranış sergilemeye fırsat tanıyan bir ağ iç durumu ortaya çıkarmıştır. RNN'ler kendi giriş belleğini ileri beslemeli sinir ağlarının aksine, girdilerin rastgele dizilerini işlemek için kullanabilmektedir. Sıralı

bilgilerin kullanılması yinelenen sinir ağlarında temel düşünceyi oluşturur. Görüntü tabanlı verilerde tüm girdi ve çıktılar birbirinden bağımsız olduğu varsayılmaktadır. Bir dizinin her unsuru için aynı görevi önceki çıktılara bağlı olarak gerçekleştirmesi söz konusudur.

RNN'lerin zaman serisi tahminlerinde tercih edilme sebebi, diğer ağlardaki kısıtlayıcı ve sabit boyutlu girişleri kabul etme özelliğidir. Bu çerçevede sabit uzunlukta çıktılar üretebilirler. Ayrıca bu modeller öğrenme sürecinde sabit hesaplama adımlarıyla işlem yapmaktadır. RNN'ler döngü oluşturabildikleri için sıralı olarak gelişen olayları birbiriyle anlamlandırmaktadır. Bu avantajından dolayı son yıllarda ağırlıklı olarak dil, modelleme, çeviri, konuşma ve tanıma gibi farklı alanlarda yaygın olarak kullanılabilmektedir. RNN'lerin dezavantajları arasında hesaplama sürecinde yavaşlık, uzun vadede önceki bilgilere ulaşmadaki zorluk ve mevcut durum için gelecekteki herhangi bir bilginin düşünülmemesi olarak sayılabilir (Soy Temür,2019:s.86).

3.4.2. Support Vector Machine (Destek Vektör Makinesi)

1963 yılında Vladimir Vapnik ve arkadaşları tarafından temelleri atılan “Destek Vektör Makineleri (DVM)” hem sınıflandırma hem de regresyon analizleri için kullanılan bir denetimli öğrenme algoritmasıdır. Sürekli bağımlı değişken ve kategorik değişkenler için kullanılabilmektedir. Veri setini ikiye ayıran doğru karar doğrusu olarak isimlendirilmektedir. Sonsuz tane karar doğrusu çizebilme imkânı mevcut olsa da önemli olan optimal yani en uygun karar doğrusunu belirlemektir. Destek vektör makinesi ikili sınıflandırmada yapısal risk minimizasyonu ilkesinin uygulaması olarak karşımıza çıkmaktadır. Destek vektör makinesi algoritmasının temel amacı ise, yüksek boyutlara sahip bir karar doğrusu tanımlayarak, iki sınıfa ait veri kümesinin uygun bir biçimde birbirinden ayrılabilmesi için pozitif ve negatif örnekler arasındaki ayrımın optimum seviyede yapılmasıdır. Bu noktalar, destek noktaları olarak adlandırılmaktadır(Güler,2019:s.52). Destek vektör makinesi; noktaları desen uzayında veya daha yüksek boyutlu bir uzayda iki ayrık yarım uzaya atayarak sınıflandırmaktadır. Temel düşüncesi, pozitif ve negatif örnekler arasındaki ayrım marjının maksimize edileceği şekilde karar yüzeyi olarak veri setini iki sınıfa ayıran iki boyutlu bir uzay olan karar doğrusu oluşturmaktır. Bir eğitim kümesi için $x_i \in \mathbb{R}^d (i= 1,2,... , N)$ girdi vektörleri ve $y_i \in \{1,-1\} (i= 1,2,... , N)$ karşılık kümesi olduğu varsayalım. Burada (1) ve (-1) iki sınıfı ifade etsin. Amaç ikili bir sınıflandırma oluşturmak veya mevcut örneklerden, daha önce görülmeyen bir örneklemden birimleri yanlış sınıflandırma ihtimali düşük olan bir karar fonksiyonu ortaya çıkarmaktır.

Giriş vektörlerini $x_i \in \mathbb{R}^d$ yüksek boyutlu bir özellik alanına $\Phi(x_i) \in \mathcal{H}$ haritalar ile sınır çizgileri eşit uzaklıkta çizilen bir karar doğrusu oluşturur. Karar doğrusu ile $\mathcal{H}$ uzayındaki her bir sınıfın en yakın veri noktaları arasındaki mesafe maksimize edilir. Φ Haritalaması $\mathcal{H}$ iç çarpım uzayında tanımlanan $K(x_i, x_j)$ çekirdek fonksiyonları tarafından gerçekleştirilir. Elde edilen sınıflandırıcı α_i katsayılarını belirlemek için eşitlik 1 de verilen karar fonksiyonu ve kuadratik programlama problemine dayalıdır.

$$f(x) = \text{sgn} \left(\sum_i^N = 1 (y_i \alpha_i \cdot K(x, x_i) + b) \right) \quad (1)$$

Kuadratik problem ise eşitlik 2' de gibi tanımlanır:

$$\text{Max.} \sum_i^N = \alpha_i - \frac{1}{2} \sum_i^N = 1 \quad \sum_j^N = 1 \quad \alpha_i \alpha_j * y_i y_j * K(x, x_i), 0 \leq \alpha_i \leq c \quad (2)$$

$$\sum_i^N = 1 \quad \alpha_i y_i = 0 \quad i = 1, 2, \dots, N$$

Burada c , marjın ile sınıflandırma hataları arasındaki dengeyi kontrol eden düzenleme parametresidir. DVM'de yaygın kullanılan polinomial eşitlik 3'de ve radyal tabanlı fonksiyon eşitlik 4'te olmak üzere iki adet çekirdek fonksiyon vardır.

$$\text{Polinom fonksiyon: } K(x_i, x_j) = (x_i \cdot x_j + 1)^d \quad (3)$$

$$\text{Radyal tabanlı fonksiyon: } K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (4)$$

Tahminin başarısı çekirdek fonksiyonu seçimine bağlıdır. İlgili fonksiyonların parametrelerinin belirlenebilmesi belirli bir metot olmamakla birlikte denemeler yapılarak belirlenir (Pabuçcu, 2009: s.184).

Avantajları:

- Yüksek boyutlu uzaylarda etkilidirler.
- Boyut sayısının, örneklem sayısından daha fazla olduğu durumlarda etkilidirler.
- Karar fonksiyonunda bir takım eğitim noktaları kullanılır ("support vectors"). Dolayısıyla bellek verimli bir şekilde kullanılmış olur.

- Çok yönlü: Karar fonksiyonu için çok farklı çekirdek fonksiyonları (Kernel functions) kullanılabilir.

Destek Vektör Makineleri, veri setinin doğrusal olarak ayrılabilme ve ayrılamama durumuna göre ikiye ayrılmaktadır. Bunlar doğrusal destek vektör makineleri ve doğrusal olmayan destek vektör makineleri olarak adlandırılır. Doğrusal destek vektör makineleriyle sınıflandırmada, iki sınıfa ait örneklerin doğrusal olarak dağıldığını varsayalım. Bu durumda bu iki sınıfın, eğitim verisi kullanılarak elde edilen bir karar fonksiyonu yardımıyla birbirinden ayrılması amaçlanır.

Doğrusal olmayan bir veri kümesinde DVM'ler doğrusal bir hiper-düzlem çizemez. Bu nedenle çekirdek numarası olarak adlandırılan *kerneltrick*'ler kullanılır. Çekirdek yöntemi, doğrusal olmayan verilerde makine öğrenimini yüksek oranda arttırmaktadır (Çimen,2020:s.62). En çok kullanılan çekirdek yöntemleri Polynomial Kernel ve Gaussian RBF (Radial Basis Function) Kernel'dir.

3.4.3. IBk (K-nearestneighbor – K-en yakın komşuluk)

K En Yakın Komşu algoritması 1950'lerin başında ortaya çıkmıştır. Bu yöntem, büyük eğitim setlerinde öğrenme işleminde zaman kaybedilmesine yol açmaktadır. Bu sebeple, bilgi işlem gücü kullanılabilir hale dönüşene kadar popülerlik kazanmamıştır. 1960'lardan sonra hazırlanan minimum uzaklık sınıflayıcı üzerine çalışmalarla geliştirilerek 1967'de T. Cover ve P. Hart'ın sunumunu gerçekleştirdiği Yakın Komşular Örüntü Sınıflama çalışmalarıyla netliğe kavuşmuştur (Dilki ve Deniz Başar, 2020:s.227).

Denetimli öğrenme yöntemlerinden biri olan En Yakın Komşu algoritması sınıflama ve regresyonda kullanılabilen çok yönlü bir algoritmadır. Basit bir ifadeyle, sınıfı bilinmeyen veri, eğitim setindeki diğer verilerle karşılaştırılarak uzaklık ölçümü gerçekleştirildikten sonra, hesaplanan uzaklığa göre henüz bir sınıfa atanamamış veriye optimal sınıf bulunur.

Örnek tabanlı olup aynı mesafe ölçüsünü kullanan K-en yakın komşular sınıflandırma algoritmasıdır. Hem sınıflandırma hem de regresyon problemleri için kullanılır. Çapraz doğrulamaya dayalı olarak uygun k değerini seçebilir. Mesafe ağırlıklandırması da yapabilir. Denetimli sınıflandırma algoritmalarından k en yakın komşu algoritmasında sınıflandırma işlemi sırasında kaç adet k değerine bakılarak sınıflandırma yapılacağına karar verilir. Gelen değerlerin sınıflandırmasında k değerine bakılarak o değere en yakın k adet değer karşılaştırılır ve en çok sınıfı olan değer sınıfı olmuş olur.

K en yakın nokta hesaplaması sırasında çeşitli uzaklık hesaplama fonksiyonlarından faydalanılır. Aşağıdaki fonksiyonlar bunların başlıcalarıdır:

1. Manhattan Hesaplaması
2. Minkowski Hesaplaması
3. Öklid Hesaplaması

Ağırlık belirleme yöntemi ile $W=1/(d*d)$ ile her uzaklığın ağırlığı bulunarak seçilen k değerine göre sınıf karşılaştırması yapılır(Kırelli,2013:s.24).

3.4.4. Random Forest (Rastgele Orman)

Rastgele orman sınıflandırma ve regresyon problemlerinde kullanılan, karar ağacı algoritmalarından oluşan denetimli bir öğrenme algoritmasıdır. Karar ağaçları rastgele bir orman algoritmasının yapı taşlarıdır. Rastgele orman algoritması tarafından oluşturulan orman, torbalama veya önyükleme toplama yoluyla eğitilir. Torbalama, tek bir örnek yerine farklı eğitim veri örneklerinin kullanılmasını içerir(Sarıkaya,2021:s.8). Davranış ve sonuçları tahmin eder ve daha çok sınıflandırma problemleri için kullanılır. Rastgele orman algoritması, veri örnekleri üzerinde karar ağaçları oluşturur. Regresyon hesaplamak için tek tek karar ağaçlarının tahminlerinin ortalaması alınır. Karar ağaçlarının tahminlerine dayanarak sonuç belirlenir. Ağaç sayılarının artması ile sonucun kesinliği artar. Rastgele oluşan bir orman, karar ağacı algoritmalarının sınırlılıklarını ortadan kaldırır. Hassasiyeti artırır ve tahminler üretir.

Rastgele orman algoritması öncelikle veri kümesinden rastgele örnekler seçer ve seçilen her bir örnek için bir karar ağacı oluşturur. Oluşturulan her bir karar ağacından bir tahmin sonucu üretir. Daha sonra tahmin edilen her sonuç için ortalama hesaplanır ve en çok oylanan tahmin sonucunu seçecektir.

3.4.5. Random Comitte (Rastgele Komite)

Veri içerisinde rastgele seçilen temel sınıflandırıcılar komitesini oluşturmak için kullanılan bir sınıflandırıcıdır. Aynı verilere dayalı olarak seçilen her temel sınıflandırıcı, farklı bir rasgele sayı çekirdeği kullanılarak oluşturulur. Bireysel temel sınıflandırıcılar ile üretilen tahminlerin düz bir ortalaması olarak nihai tahmin hesaplanır (Kara ve ark.,2021).

3.4.6. KStar (Kyıldız -K*)

K*, regresyon için geliştirilmiş örnek tabanlı bir sınıflandırıcıdır, yani bir test örneğinin sınıfı, bazı benzerlik işlevleriyle belirlendiği üzere, ona benzer eğitim örneklerinin sınıfına dayanır. Entropi tabanlı bir uzaklık ölçütü işlevi kullanması açısından diğer örnek tabanlı öğrenicilerden farklıdır. Veri madenciliğinde kullanılan öğrenme algoritmalarından biri olan bu yöntemde amaç, test veri setinde yer alan özniteliği belirli olmayan bir örneğin, veri tabanı içerisinde daha önceden kategorize edilen fakat ortaya çıkmamış eğitim veri setindeki örnekler ile karşılaştırılması esasına göre sınıflandırma gerçekleştirmektir (Vursavuş ve Üçkardeş,2015:s.302). Kstar ve random forest algoritmalarının en iyi sınıflandırıcı algoritmalar oldukları bilinmektedir. Bunun yanında en az hata oranına sahip algoritmalar içerisinde sırasıyla Random Forest ve Kstar sıralanabilmektedir.

3.4.7. Regression By Discretization (Ayrıklaştırma ile Regresyon)

Regresyon problemleri üzerinde kullanılacak sınıflandırıcıları uyarlayan bir yöntemdir. Nominal bir değer elde etmek için bağımlı değişken üzerinde ayrıklaştırma ile bir sınıf oluşturma işlevidir. Veri ayrıklaştırma, veri ön işleme yöntemlerinden birisidir. Sürekli verinin kesikli veriye dönüştürülmesini ifade etmektedir. Veri ayrıklaştırmayı Jin ve diğ. (2007) en az veri kaybı ile sürekli verinin sonlu komşu aralıklar şekline dönüştürülmesi olarak tanımlarken Das ve Vyas (2010) benzer veya yakın nitelikler taşıyan sürekli verinin grup veya aralıklara dönüştürülmesi olarak açıklamaktadır. Veri ayrıklaştırma işlemi sonrasında sürekli veri azaltılarak veri kümesi en iyi temsil edilebilecek şekilde özetlenmektedir. Veri madenciliği sonrası ise elde edilen bilgilerin kullanımı kolaylaştırarak bu bilgilerin daha anlamlı hale gelmesi sağlanmıştır(Koçoğlu,2012:s.21). Veri ayrıklaştırma işlemi için farklı yöntemleri temel alan yöntemler bulunmaktadır. Bunlar:

- Chi Merge Ayrıklaştırma Yöntemi
- Chi2 Ayrıklaştırma Yöntemi
- Eşit Genişlikli Ayrıklaştırma Yöntemi
- Eşit Frekanslı Ayrıklaştırma Yöntemi
- IRD Ayrıklaştırma Yöntemi
- ID3 Ayrıklaştırma Yöntemi
- CADD Ayrıklaştırma Yöntemi
- CAIM Ayrıklaştırma Yöntemi

3.4.8. Decision Table(Karar Ağacı)

Karar ağaçları metodu sınıflandırma ve tahmin için kullanılan bir yöntem olup sıralı kümelerden oluşur. Her sınıf için karar kuralları oluşturabilir. Karar ağacı oluşturulurken elde var olan verilerin bir bölümü eğitim için kullanılır ve karar ağacı bu yöntemle oluşturulur. Verilerin belli bir bölümü ise meydana getirilen karar ağacını test etmek için kullanılır. Karar vermede daha anlaşılır olma potansiyelini if-then komutları ile karar tablosu oluşturarak sağlamaktadır. Karar ağaçları, öngöründe bulunması ve tanımlayıcı olması gibi bazı belirleyici niteliklere sahiptir. Veri kümesinde tanımlanan öz nitelik ile karar tablosunu oluşturur. Çıktı öz nitelik ile aynı formda oluşur (Mete,2019:s.40). Eğitim ve test verisinin örneklenmesi için çapraz geçerlilik kullanılır. Çapraz geçerlilik veri kümesinin verilen oran dâhilinde bölümlenerek aynı oran dahilinde test edilmesidir. Değerlendirme için çapraz doğrulama kullanabilir. Karar ağacında mevcut olan düğümler, ağaç içinde birbirleri ile olan ilişkilerine göre ebeveyn ve çocuk olarak adlandırılır. Öğrenme modeli düğümlere uygulanan bölme işlemi ile gerçekleşir. Bölme işlemi kökten yapraklara doğru ilerlediği için tümden gelim söz konusudur. Karar ağacı öğrenme modeli verinin belli bir bölümüne uygulanır ve kalan veri test verisi olarak kullanılır. Eğitilen model test verisi üzerinde uygulanır ve farklılık oranına bakılır. Eğitim ve test verisinin örneklenmesi için çapraz geçerlilik kullanılır. Çapraz geçerlilik veri kümesinin verilen oran dâhilinde bölümlenerek aynı oran dâhilinde test edilmesidir.

Sınıflandırmada sık kullanılan modelleme algoritmalarından karar ağacının tercih edilmesindeki en önemli etkenler;

1. Maliyetinin düşük olması,
2. Yorumlamanın kolay olması ve veri tabanına kolayca entegre edilebilmesi,
3. Güvenilirliğinin yüksek olması

Karar ağacını meydana getirebilmek için eğitim kümesinde örnekleri en iyi şekilde açıklayan değişken bulunur. Seçilen dalın üzerindeki örneklerden yeni bir değişken bulunur ve yeni dallar meydana çıkar. Örnekleri ayırt edecek başka bir ayırtaç kalmadığında yeni dallanma bitmiş olur. Ayırt edici olacak, ağaç kökü hesaplamasıyla elde edilir. Değişkenlerden hangi nitelik en fazla dala sahip ise değişken kök olarak nitelik belirlenir (Kırelli, 2016:s.19).

3.5. Derin Öğrenme

Beynin sinir yapısından ve yapay sinir ağları işlevlerinden esinlenen makine öğrenme algoritması olan derin öğrenmede, biyolojik nöronlarda olduğu gibi yapay nöronlar giriş sinyallerini almaktadır ve bu sinyaller toplanarak işlenmiş olan çıkışlara iletilmektedir. Son yıllarda küreselleşmenin etkisiyle donanım ve bilişim teknolojisinin gelişmesiyle dijital ortamda her an milyonlarca veri büyük boyutlarda üretilerek depolanmaktadır. Eğitim için kullanılan veriler çoğaldıkça, klasik makine öğrenme algoritmalarının büyük verileri karşılamadaki performansı yetersiz kalmıştır. Makine öğrenme algoritmalarında hızlı gelişmeler gerçekleşmektedir; günümüzde internetin yaygın şekilde kullanılması ile verilere ulaşma kolaylaşmıştır. Sonuca hızlı ulaşmanın yanı sıra, ulaşılan sonucun performans değerinin yüksek olması son derece önemlidir. Derin öğrenme yöntemleri söz konusu verileri etkin bir şekilde kullanarak, verimli sonuçlara ulaşmayı kolaylaştırmıştır. Derin öğrenme sınıflandırma niteliklerini ve görevlerini doğrudan veriden gerçekleştirmeyi öğrenmesi yönüyle diğer alanlarda ayrılmaktadır. Derin öğrenme modelleri, farklı ve çok sayıda sinir ağı kullanarak insan düzeyinde öğrenme ve bu seviyeyi aşan bir performans göstermektedir (Şişmanoğlu ve ark.,2020:s.473). Geleneksel makine öğrenme metotları ve derin öğrenme yöntemleri karşılaştırıldığında geleneksel yöntemler veri işleme noktasında sınırlı kalmaktadır. Derin öğrenme yöntemleri ise çoklu işlem katmanlarının hesaplamalarında kullanılmaktadır (Turan, 2019:s.17).

Derin sinir ağları ise, klasik yapay sinir ağları yerine, çok fazla katman ve nörondan oluşan bir yapıdadır. Bu yapıya sahip ardışık katman bir önceki katmandaki çıktıyı girdi olarak kullanmaktadır. Derin öğrenme kullanıldığı alana göre farklı modellerde geliştirilmektedir. Derin öğrenme ağları probleme uygun öznitelikleri çıkarmakta ve çıkarılan öznitelikler ağın öğrenmesiyle ortaya çıkmaktadır. Derin öğrenme ağları kendine sunulan bilgiyi kullanmaz. Bunun yerine hangi bilgiden öğreneceğine kendi karar vermektedir. Bu yönüyle klasik yöntemlere oranla daha başarılı sonuçlar üretebilmektedir ve zaman içerisinde farklı problemlerin çözümleri içinde farklı yapıda derin sinir ağları geliştirilmiştir. Bunlar çok katmanlı derin yapay sinir ağları, tekrarlamalı sinir ağları, uzun kısa vadeli hafıza, sınırlı Boltzmann makinesi, otomatik kodlayıcı, konvolüsyonel derin sinir ağları literatürde geliştirilen derin sinir ağları modelleridir (Sezer, 2015:s.45).

Derin yineleme denklemleri vasıtasıyla karmaşık zamansal dinamikleri öğrenebilmektedir. İnsanlar bilgiyi işlerken, daha önceden öğrendikleri bilgileri, tecrübe ettikleri olayları, yeni durum ile analiz edip öyle tepki vermektedirler. RNN'ler pek çok görevde başarılı olduklarını kanıtlamıştır, ancak kısmen tekrar eden ağırlık bir çok katmanı (zaman karşılık gelen) boyunca geçişlerin neden olabileceği yok olma ve patlama eğilimlerine bağlı olarak uzun vadeli dinamikleri öğrenmede sorunlar çıkarmaktadır.

3.5.1. Uzun Kısa Süreli Bellek (LSTM)

Uzun kısa süreli bellek tekrarlayan sinir ağlarının alt dalı olarak RNN'deki sorunların ortadan kaldırılmasına yardımcı olmak için geliştirilmiş bir bellek türüdür. Uzun kısa süreli bellek yapısında giriş kapısı, çıkış kapısı ve unut kapısı olmak üzere üç kapı bulunmaktadır. Bu kapılar sırasıyla; yazma, okuma ve sıfırlamadır. Hücre durumlarındaki değişiklikler bu unsurlar tarafından denetlenmektedir. Hücrelere iletilebilecek bilgilerin seçimi, bu bilgilerin çıktı olarak nasıl kullanılacağı ve kullanılabilecek miktar gibi seçimlerin gerçekleşmesini sağlayan kapı mekanizmalarını sisteme dâhil ederek RNN'deki sorunların çözülmesine olanak tanımaktadır. Giriş kapısının fonksiyonu belleğe eklenecek bilgileri kontrol etmektir. Unut kapısı, eski bilgilerin hangi oranda yeni verilere taşınacağını kontrol etmek, çıkış kapısı ise bellekte bulunan bilgilerin ne kadarının çıktı aşamasında kullanılacağını kontrol etmektir (Sakarya ve Yılmaz, 2019: s.115).

Diğer sinir ağlarında olduğu gibi uzun-kısa süreli bellek hücresinin de “ileri” ve “geri” olmak üzere iki fazı bulunmaktadır. İleri faz aşamalarında uzun-kısa süreli bellek için çıkışlar hesaplanır. Daha sonra ileri faz çıkışları ile hatalar hesaplanır. Ağırlık değerlerini güncellemek için ise geri gaz gereklidir.

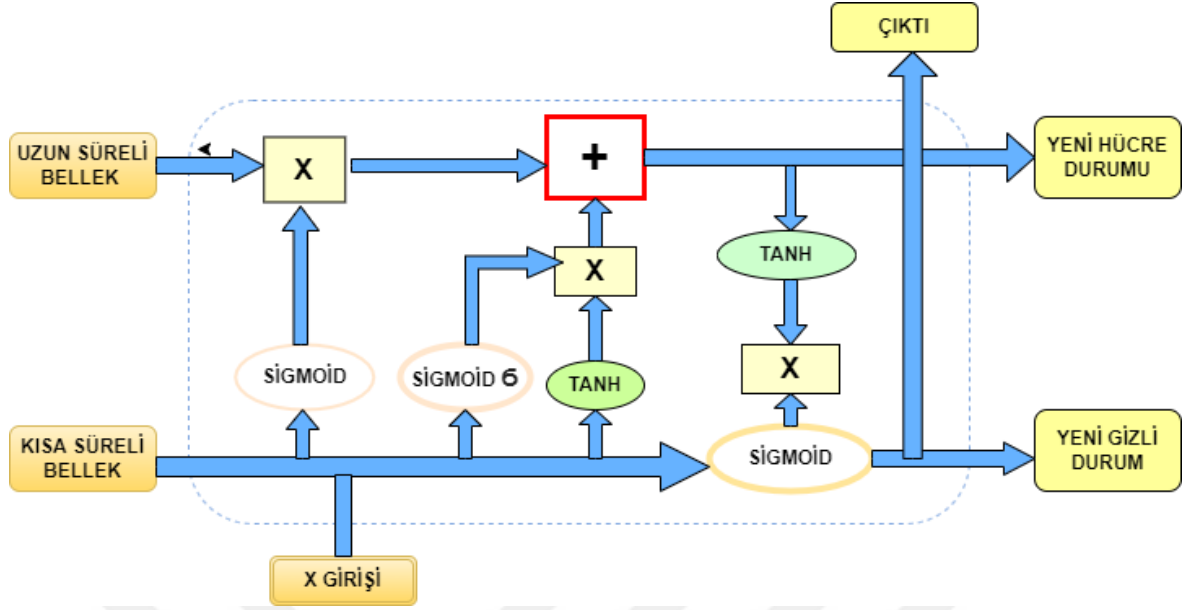
İleri faz, giriş ve önsel verileri uzun-kısa süreli bellek hücresi içerisinde işlemlerden geçirerek çıkış üretmektedir. Giriş verisi olarak uzun-kısa süreli bellek girişine gelen veri x_t ve bir önceki çıkışı ifade eden O_{t-1} birleştirilir. Bu birleşmede bir önceki verilerde kullanılmaktadır. Veri sıralaması bu açıdan önemlidir. Birleşmenin sonucunda oluşan veri Unutma Kapısı içerisindeki Unutma Kapısı Ağırlıkları (W_f) ile çarpılır. Daha sonra sonucun üzerine bu kısma özgü Unutma Kapısı Sapma Değerleri (b_f) eklenir. En sonda sigmoid fonksiyonundan geçilerek unutma kapısı çıkışına ulaşılır. Bütün işlemlerin tamamlanmasının arkasından tek bir uzun-kısa süreli bellek hücresi için Unutma Kapısı

Alanı çıkışı elde edilir. Söz konusu çıkış ilerleyen adımlardaki çıkışlarla birleştirilerek sonuca gidilir (Staudemeyer ve Rothstein Morris, 2019: 40).

Unutma kapısı çıkışının bulunmasının ardından, “*Giriş Kapısı*” ve “*Aktivasyon Kapısı*” çıkışları hesaplanır. Bu çıkış hesaplanırken “*Unutma Kapısı*” alanındaki x_t ve O_{t-1} birleştirilir. Giriş kapısı hesaplamaları için birleştirilen girişler, Giriş Kapısı Ağırlıkları ile çarpılır ve bu değere Giriş Kapısı Sapma Değerleri (b_i) eklenir. Bu bölümde oluşan değer sigmoid fonksiyonuna tabi tutularak Giriş Kapısı çıkışları oluşur. İkinci bölümde birleştirilen x_t ve O_{t-1} değerleri önce Aktivasyon Kapısı Ağırlıkları (W_a) ile çarpılır, sonucun üzerine Aktivasyon Kapısı Sapma (b_a) Değerleri eklenir. Oluşan sonuç tanh fonksiyonuna dahil edilerek Aktivasyon Kapısı çıkışları oluşur. Son olarak, Giriş Kapısı çıkışları ve Aktivasyon Kapısı çıkışları çarpılarak bu alanın asıl çıktısı elde edilir.

Sıradaki adımda Giriş Kapısı, Aktivasyon Kapısı ve Unutma Kapısı çıkışları kullanılarak Bağlam Çıkışları değerleri elde edilir. Bu değerler bir sonraki uzun-kısa süreli bellek hücresi için kullanılacaktır ve amacı veri sırası arasında bağlantı kurarak sıralamayı gözetmektir. Son olarak uzun kısa süreli bellek çıkışları hesaplanmalıdır. Çıkışları hesaplamak için bir önceki çıkışlar ile giriş verisi birleştirilerek sigmoid fonksiyonundan geçilir ve bu değer bir önceki adımda elde edilen bağlam çıkışlarını tanh fonksiyonundan geçirerek çarpılır. Elde edilen değer uzun-kısa süreli bellek hücresinin çıkışıdır.

Birden fazla çıkış olması halinde bağlam çıkışları ve hücre çıkışları bağlanarak uzun-kısa süreli bellek ağ yapısı oluşturulur. Tüm uzun-kısa süreli bellek hücreleri için çıkışlar hesaplandıktan sonra güncelleme değerleri için geri faz başlamaktadır. İleri fazda hesaplamalar en soldaki uzun-kısa süreli bellek hücresinden başlayarak en sağdaki hücrenin çıkışı elde edilene kadar sürer. Bundan sonra geri faz başlar. Geri faz ise ters yönde ilerletilerek en sağdan en sola doğru yani çıkıştan girişe doğru yönelir.



Şekil 6. LSTM (Unut Kapısı ve Çıkış Kapısı)

Bellek birimlerini bir araya getirerek bir çözüm sağlayan belirli bir tekrarlayan ağ biçimidir. Bu, ağın önceki gizli durumları ne zaman unutması gerektiğini ve yeni bilgiler verildiğinde gizli durumları ne zaman güncelleyeceğini öğrenmesine olanak tanır. Bir LSTM modeli, önceki dönemlerin oynaklık modellerini ezberleyerek geçmişe ait verileri kullanarak tahminleri potansiyel olarak artırabilir. LSTM modeli, eğitilecek daha büyük parametreler setinin maliyeti ile verilerdeki zamansal ilişkilerin tanımlanmasını otomatikleştirebilir. LSTM modellerinin sayısız finans uygulama alanları mevcuttur. Bunlar, varlık geri dönüş hareketlerinin altında yatan ve uzun bellek etkileri yakalayabilen yeni bir değişkenlik sınıfı modeli sağlarlar.

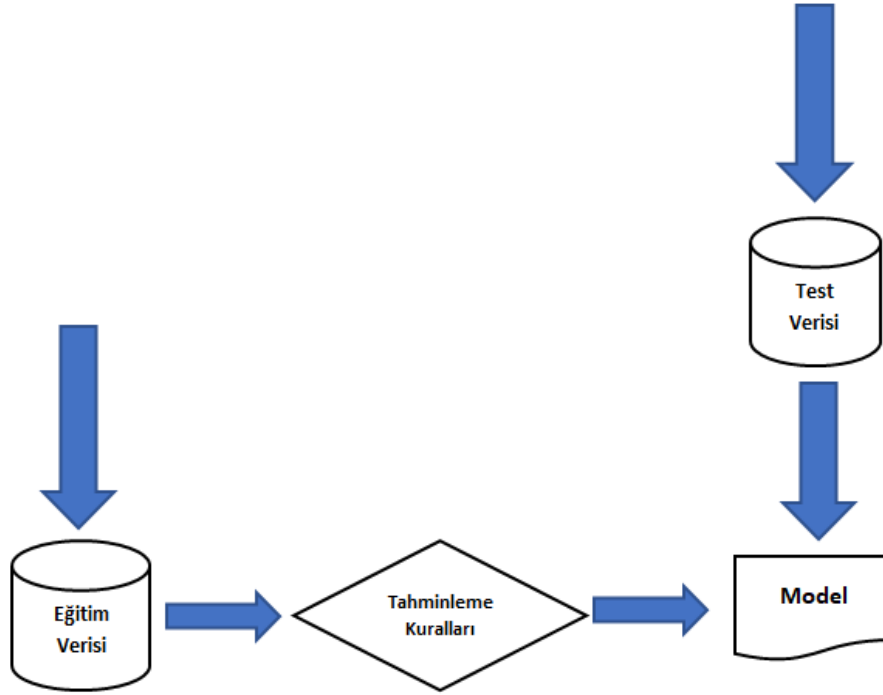
Öğrenme sürecinin verimli bir şekilde sonuçlanabilmesi amacıyla derin öğrenme uygulamalarında hata fonksiyonunun mutlak minimum değerinin hesaplanması gerekmektedir. Söz konusu süreç, optimizasyon yöntemleriyle gerçekleştirilmektedir. Optimizasyon, ağın ürettiği çıkış değeri ile gerçek değer arasındaki farkı düşürmek ayrıca ortaya çıkabilecek hatayı en düşük seviyeye indirmek için kullanılır. Yapay sinir ağlarının optimizasyonun sağlanabilmesi için en yaygın kullanılan yöntemlerden biri gradyan inişidir. Veri setinin büyüklüğüne göre üç adet gradyan inişi yöntemi (Batch Gradient Descent, Stochastic Gradient Descent, Mini –Batch Gradient Descent) bulunmaktadır. Gradyan inişi yöntemini temel alan çeşitli algoritmalar (Rmsprop, Adagrad, Adam, Nadam) mevcuttur. Optimizasyon algoritmalarında öğrenme katsayısının ayarlanması

modelin geliştirilmesinde önemli bir yere sahiptir. Fakat her algoritma ile modeldeki öğrenme katsayısını tam olarak ayarlayabilmek mümkün değildir. Bu sorunu çözebilmek için gradyan yöntemlerinin farklı çeşitleri kullanılabilmektedir.

Adam düşük dereceli momentlerin uyarlanabilir tahminlerine dayanan, stokastik amaç fonksiyonlarının birinci dereceden gradyan tabanlı bir algoritmasıdır. Adam, makine öğrenmesini birçok uygulayıcısı tarafından kullanılan teknolojisi optimizasyon algoritmalarından birisidir. İkinci an ile normalize edilen ilk an güncellemenin yönünü vermektedir. Adam optimizasyonu yönteminde, Rmsprop'ta gerçekleştiği gibi geçmiş eğitimlerin karelerinin üssel olarak ağırlıklandırılmış ortalamalarının (V_t) kullanılmasıdır. Ayrıca, momentum değişikliklerini (m_t) de önbellekte saklar. Rmsprop ve momentumu birleştirir. Varsayılan değer β_1 için 0.9; β_2 için 0.999 ve ϵ için 10^{-8} olarak belirtilmiştir (Ser ve Bati, 2019: s.410).

4. DEĞERLENDİRME ÖLÇÜTLERİ

Değerlendirme ölçütleri; öğrenim verisi, tahminleme kuralları, model ve test verisi olarak aşağıdaki şekilde ifade edilmiştir:



Şekil 7. Değerlendirme Ölçütleri

Kaynak: Data Veri Madenciliği Veri Analizi Haldun Akpınar ISBN: 978-605-4220-81-6; 2014: 228.

4.1. Çapraz Doğrulama (Cross Validation)

Makine öğrenmesi modelinde k katmanlı çapraz doğrulama olarak yer almaktadır. Modelin başarısını veri setinden doğrulama kümeleri oluşturarak ölçmektedir. Veri seti eğitim ve test verisi olarak uygulanan k katmanlı çapraz doğrulamaya göre kümelenir. Veriye 10 katmanlı çapraz doğrulama uyguladığında 9 parçasını modeli geliştirmek kalan 1 parçasını

da modeli test etmek için kullanır. Bu işlem 10 defa tekrarlanır. Her defasında farklı doğrulama kümesini test kümesi alıp kalan 9 kümeyi modeli geliştirmek için kullanır.

4.2. Ortalama Mutlak Hata (Mean Absolute Error- MAE)

Tahmin edilen değerlerin, gerçek değerlerden ne kadar uzakta olduğunu bulmaya çalışan değer ortalama mutlak hatayı vermektedir. Gerçek değerler ile tahmin edilen değerler arasındaki mutlak hatayı verir. Eşitlik 5'te MAE değerinin sıfıra yakın bir değer çıkması modelin tahmin yeteneğinin iyi olması sonucunu ortaya koyar (Duan ve ark., 2016). Regresyon ve zaman serisi analizlerinde sıklıkla kullanılmaktadır.

$$MAE = \frac{\sum_{i=1}^n |x - x'|}{n} \quad (5)$$

4.3. Ortalama Karesel Hata Karekökü (Root Mean Square Error- RMSE)

Ortalama karesel hata karekökü, tahmin edilen değerler ve gerçek değerler arasındaki uzaklığın bulunmasında hatanın büyüklüğünü ölçer. RMSE tahmin hatalarının standart sapmasını verir. RMSE tahmin hatalarının ne kadar yayıldığıнын bir ölçüsüdür. Eşitlik 6'da RMSE değeri hesaplanırken veri kümesinin hatalarının kareleri toplamı veri sayısına bölünür ve bu değerin karekökü alınır (Duan ve ark., 2016).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x - x')^2}{n}} \quad (6)$$

RMSE değerinde hataların karelerinin alınmasına bağlı olarak veri kümesindeki büyük hataların ortalama üzerindeki etkisi daha büyük olmakta ve bu sayede büyük hataların genel değerlendirmede etkisinin daha fazla olması sağlanmaktadır. RMSE değerinin genellikle ortalama bağımlı değişken değerinin %10'unun altında olması durumu başarılı bir tahminleme yapıldığı sonucunu ortaya koymaktadır (Lee, 2014). RMSE değerinin sıfıra yakın olması daha iyi performans gösterirken sıfır olması modelin hiç hata yapmadığını gösterir.

4.4. Bağıl Mutlak Hata (Relative Absolute Error-RAE)

Tahmin edilen hata oranı ile gerçek değerler farkının toplamını vererek, bu farkı gerçek değer ile gerçek değerlerin ortalaması arasındaki farkın toplamına böler (Kara ve ark.,2016). Eşitlik 7’de bağıl mutlak hata oranı tahmin edilen değerler ile gerçek değer arasında mutlak farkı vermektedir ve çıkan ortalama aritmetik ortalamaya bölüldüğünde ise bağıl olmaktadır. Çıkan bu hatanın yeterince küçük olup olmayacağına anlamak mümkündür (Botchkarev, 2018b, Sammut ve Webb,2010).

$$RAE = \frac{\sum_{i=1}^n (P_i - A_i)}{\sum_{i=1}^n (A_i - A_M)} \quad (7)$$

Eşitlik 7’de görüldüğü gibi, P_i = Tahmini Değer, A_i = Gerçek Değer, A_m = Gerçek Değerlerin Toplamı, n =Örnek sayısı ifade etmektedir.

5. GELİŞTİRME ORTAMLARI

5.1. Çalışmada Kullanılan IDE/Programlama Dili/Kütüphane

Makine öğrenmesi yöntemleri analizi ve değerlendirme ölçütleri için weka programı kullanılmıştır. Weka (Attribute Relationship File Format) programı Yeni Zellanda'nın Waikoto Üniversitesi tarafından java yazılım dili ile geliştirilmiştir. Makine öğrenmesi algoritmaları için yaygın olarak tercih edilmektedir. Açık kaynak kodlu olan bu uygulamada sınıflandırma (classification), kümeleme (clustering) ve ilişkilendirme (association) modelleri kullanılarak veri madenciliği işlemleri yapılmaktadır. Weka dosya yapıları ARFF, cvs, c45olarak kullanılmaktadır.

Çalışmada ayrıca python programlama dili ile LSTM için tahmin modelleri geliştirilmiştir. Yapay zekâ çalışmalarının %57'sinde kullanılan en yaygın programlama dili Phyton'dur (Kızrak, 2018). Phyton içerisinde doğal dil işleme, görüntü işleme, makine öğrenmesi, derin öğrenme ve web uygulamaları gibi çok farklı alanda kullanılabilecek çeşitli kütüphaneler yer almaktadır.

Tablo 5. Çalışmada Kullanılan Programlama Dili

Çalışmada Kullanılan IDE	Versiyon	Programlama Dili	Versiyon	Kütüphane	Versiyon
Jupyter Notebook IDE	6.3.0	Python Programlama Dili	3.10.1	Keras	2.4.3
				Tensorflow	2.7.0
				Numpy	1.19.5
				Matplotlib	3.3.4
				Pandas	1.2.4
				Plotly	4.14.3
				Scikit-learn	0.24.1

Tensorflow ve Theano'ya katman olarak yazılan ve kolay bir model geliştirmeyi hedefleyen Python kütüphanesidir. Francois Chollet tarafından hazırlanmıştır. Keras gelişmiş algoritmaları optimize ederek normalleşme rutinleriyle destekler. Ayrıca aktivasyon fonksiyonları olan en iyi seçenekleri ortaya koyar. Katman olarak Theano ve Tensorflow desteklese de, giriş verilerinin farklı boyutlarda olmasından dolayı her iki kütüphane içinde arka uçta çalışılabilmesi için Python dikkatli şekilde tasarlanmalıdır.

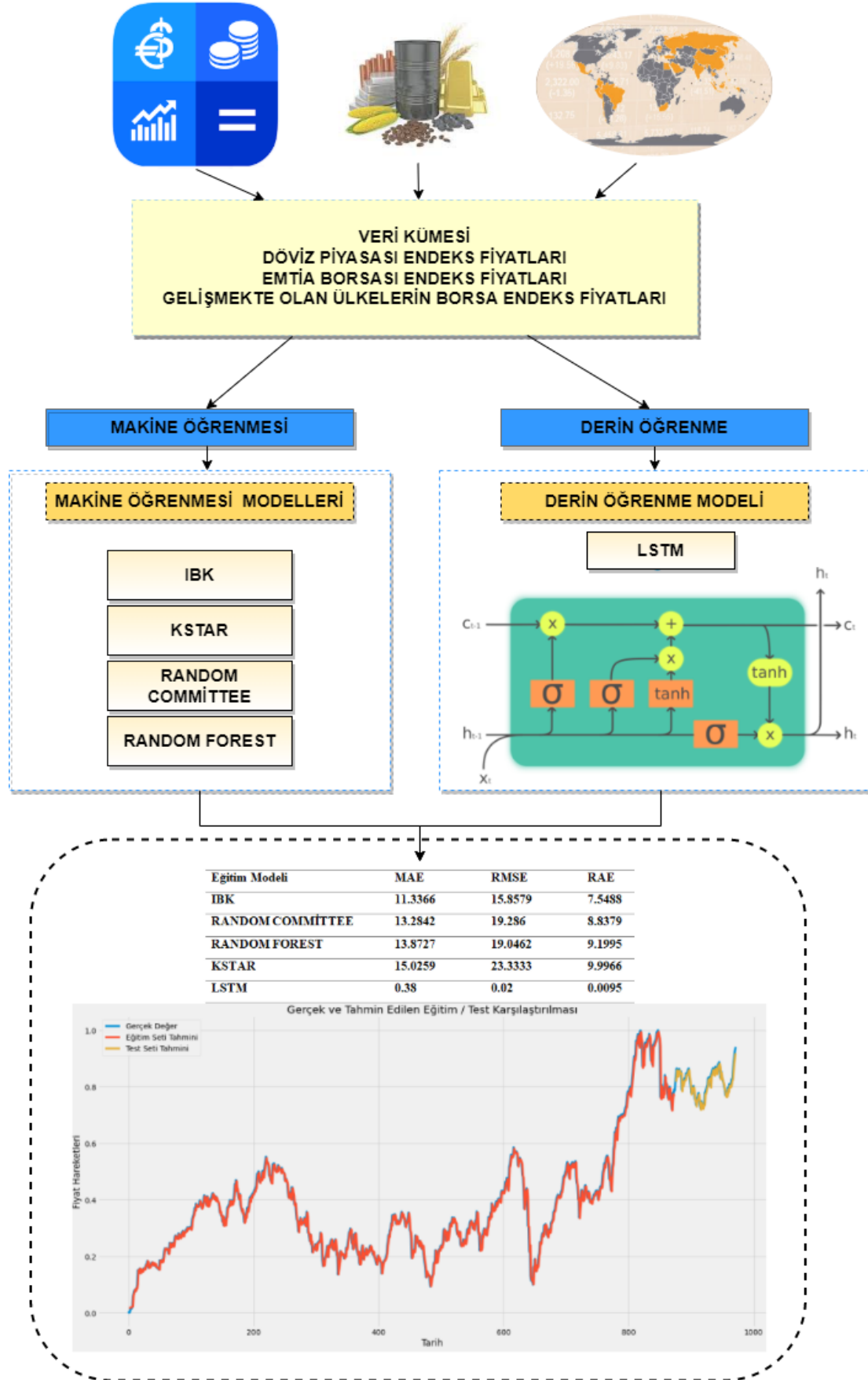
Tensorflow 2015'te Google tarafından çıkartılan açık kaynaklı bir kütüphanedir. Java, C++, Go ve Python gibi dillere destek vermektedir. Veri akış grafikleri sayesinde sayısal hesaplamalar için kullanılan bir yöntemdir. İlk defa makine öğrenimi ve derin sinir ağıları araştırmaları için geliştirilen Tensorflow, Google içerisinde bir kullanılmaktan başlayarak geniş alanlara yayılmıştır. Esnek mimarisi hesaplamayı API kullanılarak ilgili cihazlardaki bir veya daha fazla CPU veya GPU gibi birimlere dağıtmaktadır. Söz konusu işlemler düğüm olarak nitelendirilmekte ve bu düğümler arasındaki kenarlara tensör ismi verilmektedir. Tensorflow'da veri integer, float, string şekilleriyle ifade edilmemektedir. Bu değerler tensör adlı nesnede bulunmaktadır. Tensörler, çok boyutlu verileri tutan dizi yapılarıdır.

Tensorflow, ağı içerisindeki gradyanları otomatik hesap ve hata fonksiyonunu optimize eder. İlgili hesaplar bir veya daha çok GPU ile önemli derecede hızlandırılır. Tensorflow taşınabilir yapıdadır ve bu yönde gelişmekte olan bir topluluk vardır. Derin sinir ağı modelleri için geliştirme, analiz, iyileştirme gibi işlemler için Tensorboard adlı bir aracı bulunmaktadır ve bu aracı son derece kullanışlıdır (Erhandı, 2020: s. 37).

6. BULGULAR

Finansal piyasalarda geleceğe yönelik tahmin için birçok yöntem geliştirilmiştir. Son dönemde makine öğrenmesi, derin öğrenme yapay zekâ başlıkları altında birçok çalışma yapılmaktadır. Bu çalışmada da emtia, döviz fiyatları ve gelişmekte olan ülkelerin borsa endeks verileri makine öğrenmesi ve derin öğrenme modelleri Weka programı ve python yazılım dili yardımıyla incelenmiştir. Makine öğrenmesi yöntemleri ve Uzun Kısa Süreli Bellek (LSTM) yönteminin tahmin değeri karşılaştırılmıştır. BİST100 endeksinin kapalı olduğu günlere ilişkin olarak veri setlerinde ayrıştırma yapılmıştır.

Verileri tahmin etmek ve değerlendirmek için makine öğrenmesi modelleri Weka programı ile analiz edilmiştir. Makine öğrenmesi modelleri değerlendirilerek en yüksek başarı oranını veren 10 farklı makine öğrenmesi yöntemi ile 5 ve 10 fold (k katmanlı çapraz doğrulama) kullanılarak deneysel çalışmalar yapılmıştır. Makine öğrenmesi modellerinin performansında hata oranları ortalama mutlak hata (Mean Absolute Error-MAE), ortalama karesel hata karekökü (Root Mean Squarederror-RMSE) ve (Relative Absolute Error-RAE) değerlendirme ölçütleri ile incelenmiştir. Uygulanan veri setinin 14 özniteliği bulunmaktadır. Analizler için Intel(R) Core™ i5-5300U CPU@ 2.30GHz ve 8 GB belleğe sahip bilgisayar kullanılmıştır. Çalışmada aynı veri seti üzerinde makine öğrenmesi ve derin öğrenme modelleri ile analiz yapılmıştır. Çalışmada oluşturulan veri seti üzerinde yapılan farklı deneyler ile düşük hata oranına sahip modelinin belirlenmesi hedeflenmiştir.



Şekil 8. Çalışmanın Mimarisi

Makine öğrenmesi modelleriyle tahmin yapma ve performans karşılaştırmasına ilişkin tablo aşağıda yer almaktadır. Tabloda 5 ve 10 fold ile hesaplanan hata oranları yer almaktadır.

Tablo 6. Makine Öğrenmesi 5 ve 10 Çapraz Doğrulama ile Performans Karşılaştırması

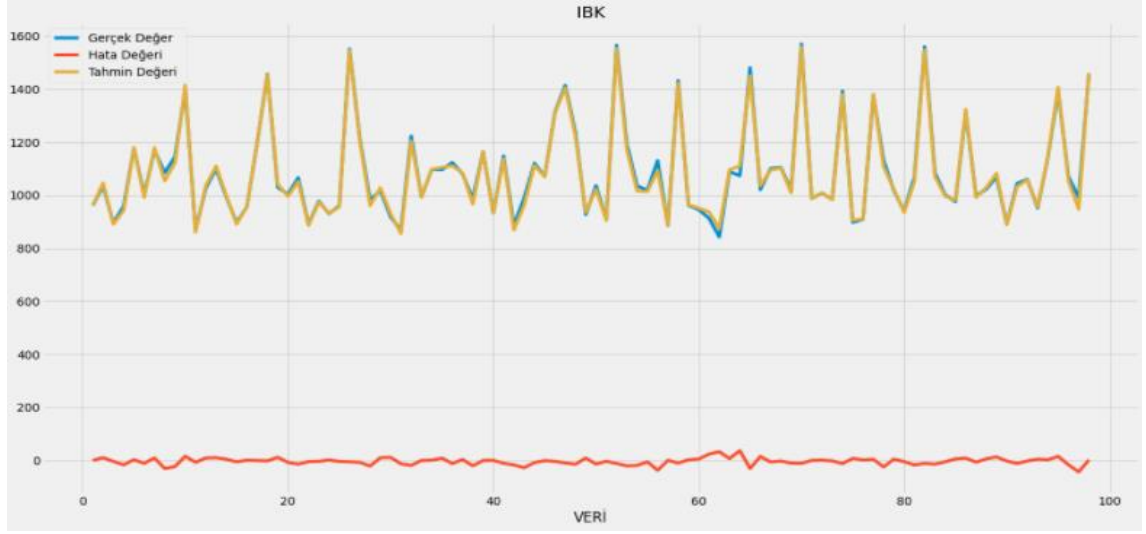
SINIF	MAE5	RMSE5	RAE5	MAE10	RMSE10	RAE10
Multilayer Perceptron	24.9645	31.6093	16.617	24.556	31.1729	16.3285
IBk	11.5213	16.1133	7.6689	11.172	15.6026	7.4288
KStar	15.0876	23.1503	10.0427	14.9643	23.5164	9.9505
Bagging	19.1072	26.5804	12.7182	18.2429	25.2856	12.1306
Random Committee	13.5368	19.9514	9.0104	13.0316	18.6206	8.6654
Randomizable Filtered Classifier	15.8639	24.3361	10.5594	15.1916	23.4192	10.1016
Random Sub Space	19.1738	26.2277	12.7625	18.4208	24.9351	12.2489
M5Rules	21.8232	30.2389	14.5261	20.3456	28.0132	13.5288
M5P	19.8036	26.9053	13.1818	19.4991	26.5143	12.9659
Random Forest	14.1467	19.5171	9.4164	13.5987	18.5753	8.9826

Tablo 6’da makine öğrenmesi algoritmalarının ile weka üzerinde yapılan analizlerin hata değerleri gösterilmiştir. Makine öğrenmesi yöntemlerine 5 fold ve 10 fold çapraz doğrulama kullanılarak yapılan tüm analizlerden 10 başarılı yöntemde detaylı olarak tabloda yer almaktadır. Algoritmaların performansları incelendiğinde en başarılı 5 öğrenme yöntemin sırası ile IBk, KStar, Random Forest, Random Comittee’dir. Sonuçların objektif değerlendirilebilmesi için yöntemlere ayrıca 10 fold çapraz geçerlilik uygulanarak analizler yapılmıştır. İki farklı çapraz geçerlilik ile yapılan hata oranlarının ortalaması alınarak sonuçların objektif olarak değerlendirilmesi sağlanmıştır.

Tablo 7. Makine Öğrenmesi Performans Karşılaştırması

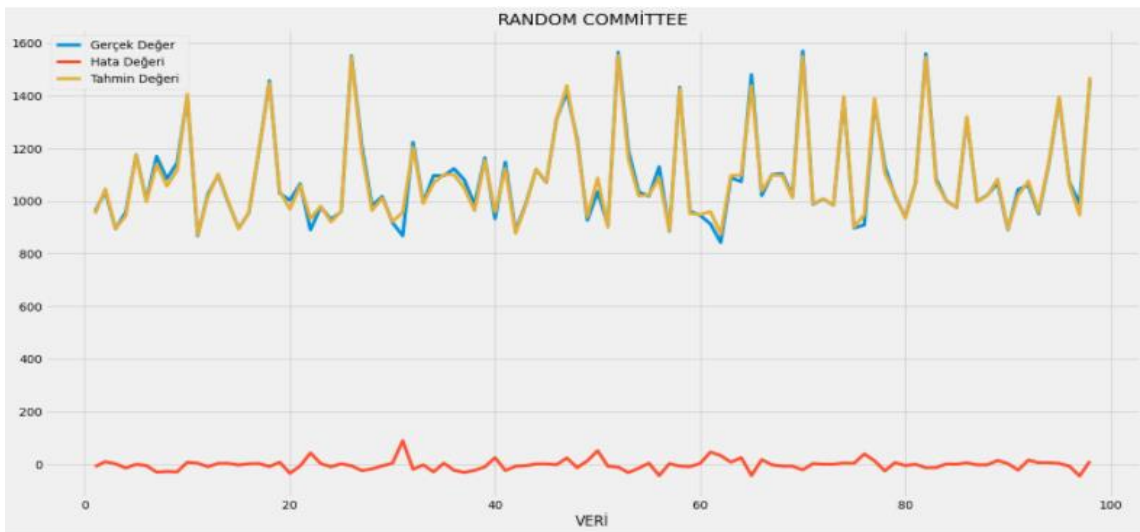
SINIF	ÇAPRAZ GEÇERLİLİK	MAE	RMSE	RAE
IBk	5 FOLD	11.5213	16.1133	7.6689
	10 FOLD	11.172	15.6026	7.4288
	ORTALAMA	11.3366	15.8579	7.5488
KStar	5 FOLD	15.0876	23.1503	10.0427
	10 FOLD	14.9643	23.5164	9.9505
	ORTALAMA	15.0259	23.3333	9.9966
Random Committee	5 FOLD	13.5368	19.9514	9.0104
	10 FOLD	13.0316	18.6206	8.6654
	ORTALAMA	13.2842	19,286	8.8379
Random Forest	5 FOLD	14.1467	19.5171	9.4164
	10 FOLD	13.5987	18.5753	8.9826
	ORTALAMA	13.8727	19.0462	9,1995

Tablo 7’de en başarılı beş yöntem ve iki farklı çapraz doğrulama uygulanması sonucunda oluşan değerler yer almaktadır. 5 ve 10 fold çapraz doğrulama uygulanan modellerin içerisinde IBk, Kstar, Random Committee, Random Forest modellerinin hata oranı diğer modellere göre daha düşük olduğu görülmektedir. Verilere en uygun parametreleri bulmak için çok sayıda deneme yapılmıştır. Denemeler sonucunda en az hata oranlarına sahip modellerin gerçek değer, tahmin değeri ve hata değerlerinin grafik olarak görselleştirilmiş hali verilmiştir.



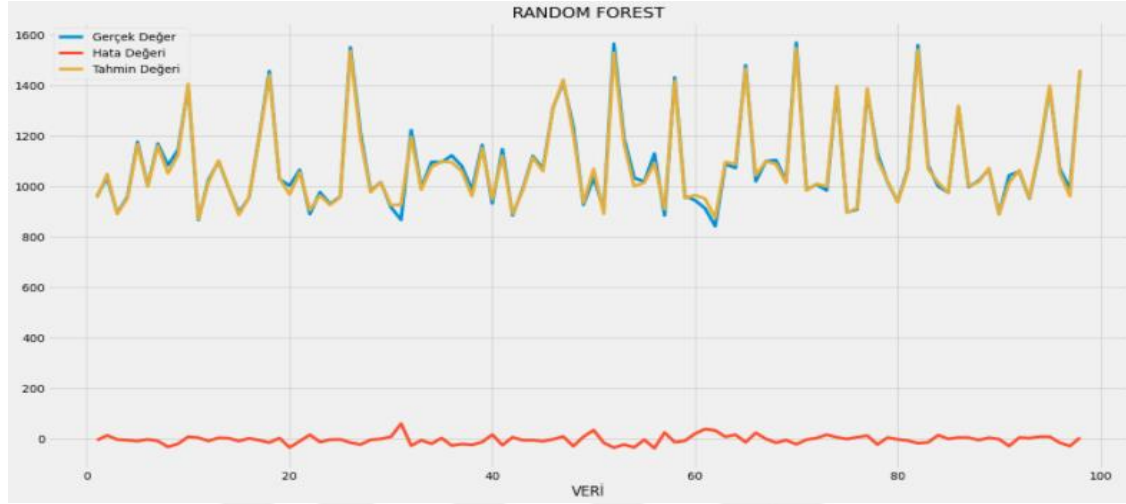
Şekil 9. IBk Tahmin Grafiği

IBk sınıflandırma yönteminde en yakın komşulara göre sınıflandırma yapılmaktadır. K en yakın komşulardaki k değeri değişken bir özellik göstermekte olup belli olmayan bir durumun ne şekilde sınıflandırılabileceğine karar vermek için kullanılmaktadır. Kendine komşu olarak kaç değer seçilebileceğini belirler. Performansı parametreler etkiler. Çalışmada farklı parametreler sonucunda oluşan yüksek performanslar baz alınmıştır. KNN= 2, batchSize=100 parametreleri ve numDecimalPlaces = 2, windowSize=0 olarak kullanılmıştır. Şekil 9’de kullanılan parametre oluşan değerler grafik olarak görselleştirilmiştir.



Şekil 10. Random Committee Tahmin Grafiği

Veri içerisinde rastgele seçilen temel sınıflandırıcılar komitesini oluşturmak için kullanılan bir sınıflandırıcıdır. BatchSize = 100 parametreleri ve numDecimalPlaces = 2, numIterations = 10, seed = 1 olarak kullanılmıştır. Şekil 10'de kullanılan parametre oluşan değerler grafik olarak görselleştirilmiştir.



Şekil 11. Random Forest Tahmin Grafiği

Rastgele orman algoritması, veri örnekleri üzerinde karar ağaçları oluşturur. Karar ağaçlarının tahminlerinin ortalaması alınarak hesaplanır. Çalışmada farklı parametreler sonucunda oluşan yüksek performanslar baz alınmıştır. BagSizePercent = 100, batchSize = 100 parametreleri ve numDecimalPlaces = 2, numIterations = 100, seed = 3 olarak kullanılmıştır. Şekil 11'de kullanılan parametre oluşan değerler grafik olarak görselleştirilmiştir.



Şekil 12. KStar Tahmin Grafiği

Kstar sınıflandırma yönteminde sınıflandırıcı bazı benzerlik özelliklerine göre ona benzeyen eğitim örneklerinin sınıfına dayanır. Burada parametre olarak batchSize = 100 ve globalBlend = 5, numDecimalPlaces = 2 olarak kullanılmıştır. Şekil 12’de kullanılan parametre oluşun değerler grafik olarak görselleştirilmiştir.

Makine öğrenmesi tahmin yöntemlerini değerlendirdiğimizde belli hata oranları görülmektedir. Literatürü incelediğimizde günümüzde popüler olarak derin öğrenme modellerinin daha başarılı sonuçlar verdiğini ve bu doğrultuda uzun-kısa süreli bellek (LSTM) ağı değerlendirilmek üzere çalışmaya eklenmiştir.

Tablo 9. Borsa Tahminlemesi İle İlgili Literatür Taraması

Yazarlar	Yıl	Yöntemler	Ölçütler	
Cao ve ark.	2005	Linear models ANN models	MAD	0,0130
			MSE	0,3437
			MAPE	0,2951
			SD	0,1247
Vamitha ve ark.	2012	Fuzzy time series The Multivariate Markov Chain model	RMSE	1,1489
Fischer ve Krauss	2012	LSTM	RMSE	0,0206
		RAF		
		DNN		
		LOG		

Shen ve ark.	2015	SAE SVR ANN	MSE MAE	0,0022 0,0061
--------------	------	-------------------	------------	------------------

Krauss ve ark.	2017	GBT RAF	RMSE	0,37
----------------	------	------------	------	------

Uzun-Kısa Süreli Bellek Modeli (Long-Short Memory-LSTM) ile tahmin yöntemi Python 3.7 programlama dilince geliştirilmiştir. Pandas, Numpy, Keras, Matplotlib, Plotly, Scikit-learn, Tensorflow kütüphaneleri kullanılmıştır. Analiz aşamasında veri seti programa okutulmuş ön işleme aşamasında veri ölçeklendirilmiştir. Ölçeklendirme (-1,1) olarak düzenlenmiştir. Model oluşturma aşamasında veri eğitim ve test verisi olarak bölünmüştür. K-Fold 10 değerine göre eğitim (%90) ve test (%10) verileri olmak üzere ikiye ayrılmıştır. Eğitim verisi LSTM ile eğitilerek model oluşturulmuştur. Tek katmalı LSTM modeli oluşturulurken katmandaki sinir sayısı 15 olarak belirlenmiş Optimizer *Adam*, hata fonksiyonu MSE olarak düzenlenmiştir. Bu parametreler dışında diğer parametreler; öğrenme için ayrılan veri kümesi (train_X), öğrenme verisi için hedef değişken (train_Y), doğrulama için ayrılan 2 veri oranı (verbose), 20 devir sayısı (epochs) modeli oluşturma esnasında tanımlanmıştır. Model oluşturulduktan sonra test verisine uyarlanarak MAE, RMSE ve RAE ile modelin hata oranları ölçülmüştür.

```
Model: "sequential_1"
```

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 15)	1020
dense_1 (Dense)	(None, 1)	16
Total params: 1,036		
Trainable params: 1,036		
Non-trainable params: 0		

Şekil 13. Kullanılan Modeller ve Parametreler

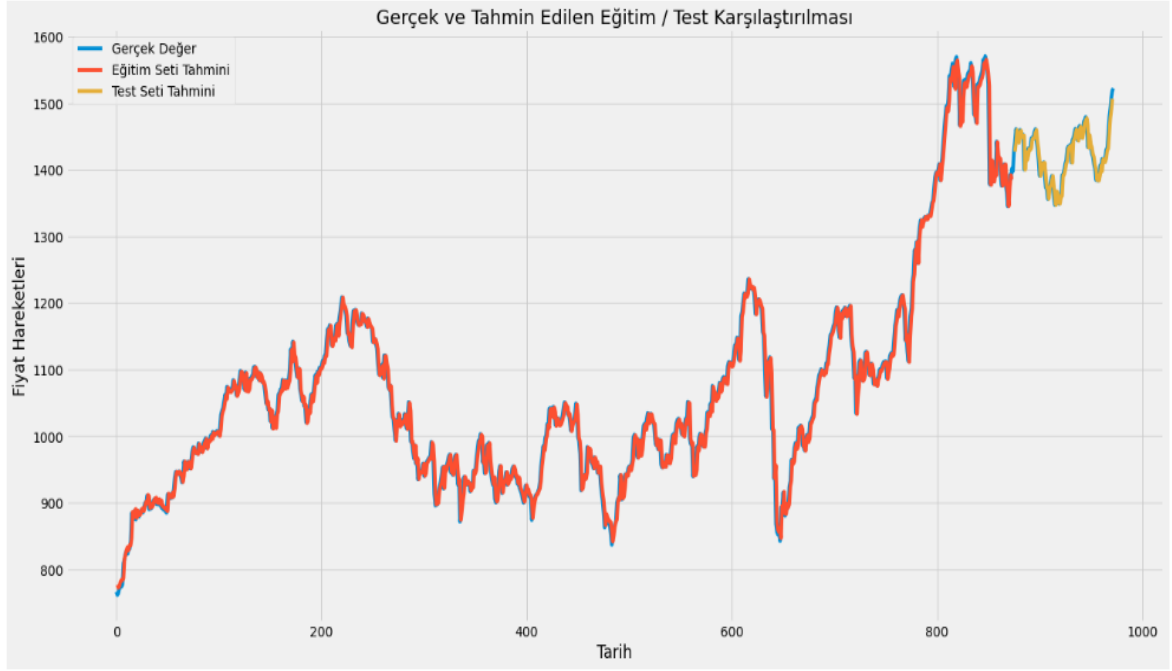
Şekil 13'te LSTM yapısı yer almaktadır. Tek katmanlı LSTM modeli ve tek bir çıkış ile düğüm kullanılmıştır.

Makine öğrenmesi ve derin öğrenme modelleri kullanılarak aynı veri kümesi üzerinde yapılan çalışmaların sonucu tablo 10' de yer almaktadır.

Tablo 8. Performans Değerleri

Eğitim Modeli	MAE	RMSE	RAE
IBK	11.3366	15.8579	7.5488
RANDOM COMMITTEE	13.2842	19.286	8.8379
RANDOM FOREST	13.8727	19.0462	9.1995
KSTAR	15.0259	23.3333	9.9966
LSTM	10.27	14.15	6.06

Tablo 10'de kullanılan makine öğrenmesi teknikleri ve LSTM modelinin ölçüm değerleri sonucu yer almaktadır. IBk modelinin MAE değeri **11.3366**, RMSE değeri **15.8579**, RAE değeri **7.5488**'dir. Random Committee modelinin MAE **13.2842** değeri RMSE değeri **19.286**, RAE değeri **8.8379**'dir. Random Forest modelinin MAE **13.8727** değeri RMSE değeri **19.0462**, RAE değeri **9,1995**'dir. KStar için MAE değeri **15.0259**, RMSE değeri **23.3333** ve RAE değeri **9.9966**'dir. LSTM modelinin MAE değer **10.27**, RMSE değeri **14.15** ve RAE değeri **6.06** 'dur. Sonuçlar incelendiğinde aynı veri seti üzerinde yapılan çalışmada sonuçlar incelendiğinde LSTM modelinin en yüksek başarıya ulaştığı görülmektedir.



Şekil 14. LSTM Tahmin Grafiği

Şekil 14’de LSTM modelinin MAE değeri **10.27**, RMSE değeri **14.15** ve RAE değeri **6.06** ile oluşan hataların grafik modeli yer almaktadır. İşlenmemiş veri(True Value) grafikte mavi renkte, eğitim verisi(Training Set Prediction) grafikte kırmızı renkte ve test verisine uygulanan eğitilmiş model(Test Set Prediction) sarı renk ile gösterilmektedir. Toplam 973 günlük kapanış verinin 875 tanesi eğitim verisi 98 tanesi test verisi olarak kullanılmıştır. Grafik ayrıca hata oranları dikkate alındığında mavi ve sarı ile oluşan trend oldukça uyumlu ve hata oranları da düşüktür.

7. SONUÇ VE ÖNERİLER

Finansal piyasalarda uygulanmak üzere, geleceğe yönelik tahmin için birçok yöntem geliştirilmiştir. Geleceği ön görmek doğru finansal kararlar vermemizi sağlamaktadır. Tahminleme için farklı yöntemlerle birçok tahmin teorisi hesaplanabilmektedir. Son dönemde makine öğrenmesi ve yapay zekâ hakkında birçok çalışma yapılmaktadır. Bu çalışmalar ile finansal tahminlere, makine öğrenmesi ve derin öğrenme modelleri ile ulaşılabilmektedir.

Makine öğrenmesi ve derin öğrenme kullanılarak yapılan çalışmada, Bist100 (Borsa İstanbul / Türkiye) ve gelişmekte olan ülkelerin borsa endeksleri ele alınmıştır. Çalışmada ayrıca Emtia borsaları, döviz piyasalarında yer alan dolar (\$), pound (£), Euro (€) ve fiyatları ve arasındaki ilişkiler Weka programı ve Python yazılım dili yardımıyla incelenmiştir. Gelişmekte olan ülkelerin borsa endeksleri, döviz kurları ve emtia borsalarına ilişkin veriler Ocak 2017 – Ekim 2021 tarihleri arasında gerçekleşen gün sonu kapanış fiyatlarıdır. Borsanın kapalı tutulduğu günler bu değerlendirmenin dışında bırakılmış olup 973 gün verisi ile analizler gerçekleştirilmiştir. Öznitelik seçiminde, piyasa ve literatürde kullanılan veriler değerlendirilerek endeks ve parite tabanlı yaklaşım uygulanmıştır. Veriler, K-fold 10 değerine göre eğitim (%90) ve test (%10) verileri olarak ikiye ayrılmıştır. Kullanılan makine öğrenmesi yöntemleri ve LSTM modeli ile hata oranları değerlendirilmiştir. Modelde kullanılan epochs, batch_size, verbose parametreleri ile oluşan hata oranları karşılaştırılarak LSTM üzerinde geliştirme gerçekleştirilmiştir. Tahmin sonucunda elde edilen ölçüm değerleri, LSTM modelinin diğer algoritmalara göre daha iyi sonuç verdiğini göstermektedir. Uygulamalar sonucunda makine öğrenmesinde en başarılı sonucu veren IBk algoritması olmuştur. LSTM için test sonuçları incelendiğinde MAE değer 10.27, RMSE değeri 14.15 ve RAE değeri 6.06 tahmin değerleri oluşurken makine öğrenmesi algoritmalarından en iyi sonucu veren IBk modeli için MAE değeri 11.3366, RMSE değeri 15.8579, RAE değeri 7.5488 tahmin değerleri görülmektedir.

Makine öğrenmesi yöntemleriyle borsa tahminlemesi yaparken 14 öznitelik üzerinden gerçekleştirilen değerlendirmeler içerisinde en düşük hata oranını IBk ve Kstar algoritmaları verirken Derin öğrenme yöntemlerinden LSTM hata oranı daha düşüktür.

Gelişmekte olan ülkelerin borsa endekslerinin tahmin edilmesine ilişkin yapay zekâ algoritmalarının kullanıldığı bu çalışma bilişim teknolojilerinin gerçekleşmesine paralel olarak gelecekte gerçekleştirilebilecek olan çalışmalara referans niteliğinde olacaktır. Farklı makine öğrenme yöntemlerinin performansının test edilebilmesi için gerekli süre ve yöntemlerin araştırılması ve geliştirilmesi, gelişmekte olan ülkelere tahminlere dayalı kısa ve uzun vadeli politikalara doğrudan yansımaktadır.

Çalışmada emtia ve döviz kurları ve gelişmekte olan ülkelerin borsa endeksleri arasındaki ilişki uzun-kısa süreli hafıza modelinin etkili bir tahmin edici olduğu ortaya çıkmıştır. Dolayısıyla bu modelin borsalar başta olmak üzere konu ile ilgili olan herkese ve bu alanda yapılacak olan bilimsel çalışmalar için katkı sağlayacağı öngörülmektedir. Özellikle ekonomik kriz dönemlerinde yapılacak olan başarılı endeks tahminleri ile yatırımcılar piyasanın geleceği hakkında öngörü sahibi olacaklardır.

Çalışmada kullanılan veri seti belli bir dönemi 2017 Ocak ve 2021 Ekim aylarını kapsamaktadır. Bu veriler farklı ülke borsalarının farklı dönemlerini ele alarak uygulamış olduğumuz model ile incelenerek gelecekte ortaya çıkabilecek değişimler tahmin edilebilir. Bu değişimler pozitif olarak ortaya çıkması yatırımcıların gerçekleştirecekleri yatırımlar konusunda daha doğru kararlara ulaşmalarına, negatif olarak ortaya çıkması halinde yatırımcıların zarar etmemeleri konusunda tedbir almalarına yardımcı olacaktır. Teknolojinin gelişmesi finans piyasaları üzerinde etkinliğinin her geçen gün artması ve insan faktörü yerine geçmesi hayalden öte bir gerçeklik olarak son yıllarda karşımıza çıkmaktadır. Bu bağlamda bu çalışmanın ortaya koymaya çalıştığı model ile bunun mümkün olduğu kanıtlanmış oldu. Uygulama sonucunda uzun kısa süreli hafıza modelinin Bist100 endeks fiyatını tahmin etmekte etkili bir yöntem olduğu sonucuna ulaşılmıştır.

Bazı algoritmalar farklı koşullarda devamlı olarak gerçek değerin altında veya üzerinde tahminler yapmaktadır. Bu algoritmaların birlikte kullanılması daha üstün performans göstermesini sağlayabilecektir. Farklı öğrenme süreleri ve farklı piyasalar üzerinde her bir algoritmanın performansının en üst seviyeye ulaşabileceği süreler araştırma konusu yapılarak farklı sonuçlara da ulaşılabilir. Çalışmanın sınırlılığı belirli tarihsel aralıktaki

(Ocak 2017- Ekim 2021) menkul deęerleri kapsamasıdır. Hedef alıřma olarak farklı finansal veri setleri ve farklı algoritmaların kullanılması ile başarı oranının arttırılması düşünölmektedir.



8. KAYNAKÇA

- Akal, M. (2013). BİST 100 endeksi ile GSPTSE ve yükselen borsalar arası etkileşim. *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler E-Dergisi*, 38, 1-20. 10 Ağustos 2021 tarihinde https://arastirmax.com/tr/system/files/dergiler/9383/makaleler/1/38/arastrmx_9383_pp_1-20_3.pdf sayfasından erişilmiştir.
- Akpınar, H. (2014). *Data veri madenciliği veri analizi*, İstanbul: Papatya Bilim.
- Alpaydın, E. (2004). *Introduction to machine learning*. The MIT Press, 3-6.
- Alpaydın, E. (2010). *Introducton to machine learning second edition*. The MIT Press, Cambridge,6.
- Ateş, Y. ve Karabatak, M. (2017). Nicel birliktelik kuralları için çoklu minimum destek değer. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 29(2), 57-65.
- Aydın, S. ve Özkul, A. E. (2015), Veri madenciliği ve Anadolu Üniversitesi Açık öğretim sisteminde bir uygulama, *Eğitim ve Öğretim Araştırmaları Dergisi*, 4(3).
- Büyüköztürk, Ş. (2012). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Akademi.
- Can, Ş. ve Gerşil, M. (2018). Manisa pamuk fiyatlarının zaman serisi analizi ve yapay sinir ağı teknikleri ile tahminlemesi ve tahmin performanslarının karşılaştırılması, *Manisa Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Yönetim ve Ekonomi Dergisi*, 25(3), 1018-1031.
- Cao, Q. Leggio, K. ve Schniederjans, M. (2005). A comparison between fama and french's model and artificial neural networks in predicting the chinese stock market, *Computerand Operations Research*, 32(10), 2499-2512.

- Çalışkan, E. S., Yazıcıoğlu. S. A., Demirci, U. & Kuş, Z. (2018). *Yapay sinir ağları, kelime vektörleri ve derin öğrenme uygulamaları*, (Yüksek Lisans Tezi), Fatih Sultan Mehmet Vakıf Üniversitesi, Mühendislik ve Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, İstanbul.
- Demirci, M. Y. ve Yabanova, İ. (2019). Model tabanlı tasarım metotları kullanılarak gerçek zamanlı bir görüntü işleme sisteminin tasarımı ve gerçekleşmesi, *Politeknik Dergisi*, 22(4), 827-838.
- Deng, L. ve Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends in Signal Processing*, 7(4), 197-387.
- Diler, A. İ. (2003). İMKB ulusal-100 endeksinin yönünün sinir ağlarıyla hata geriye yayma yöntemi ile tahmin edilmesi. *İMKB Dergisi*, 7(26), 65-81.
- Dilki, G. ve Deniz Başar, Ö. (2020). İşletmelerin finans tahmininde k – en yakın komşu algoritması üzerinden uzaklık ölçütlerinin karşılaştırılması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 19(38), 224-233.
- Ding, X et al. (2015). *Deep learning for event-driven stock prediction*. In: IJCAI.
- Duan, Y., Yisheng, L. V. ve Wang, F. Y. (2016). Travel time prediction with LSTM neural network, *In 2016 IEEE 19th international conference on intelligent transportation systems (ITSC)* (pp. 1053-1058). IEEE.
- Elmas, Ç. (2011). *Yapay zeka uygulamaları*, Ankara: Seçkin Yayıncılık.
- Erhandı, B. (2020). *Derin öğrenme ile metin özetleme*. (Yüksek Lisans Tezi), Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya.
- Fausett, L. (1994). *Fundamentals of neural network*. PrenticeHall: Hoboken
- Fertekligil, A. (2000). *Türkiye’de borsanın tarihçisi*. İstanbul: İMKB Yayınları.
- Fischer, T., Kraus, C.(2018). Deep learning with long-short term memory networks for financial market predictions. *European Journal of Operational Research*, 000,1-16.
- Gültepe, Y. (2019). Makine öğrenmesi algoritmaları ile hava kirliliği tahmini üzerine karşılaştırmalı bir değerlendirme. *Avrupa Bilim ve Teknoloji Dergisi*, 16, 8-15.

- Günay Atbaş, A. C. (2008). *Kümeleme analizinde küme sayısının belirlenmesi üzerine bir çalışma*. (Yüksek Lisans Tezi), Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Han, J. ve Kamber, M. (2000). *Data mining: concepts and techniques*. San Francisco: Morgan Kaufmann.
- Kara, Yakup, Boyacioglu, Melek A., Baykan, Ömer K. (2011) Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, 38, 5311 - 5319.
- Kara, Ş.E. ve Şamlı R. (2021). Yazılım Projelerinin Maliyet Tahmini için Weka’da Makine Öğrenmesi Algoritmalarının Karşılaştırılmalı Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, 415-426.
- Karslı, M. ve Karaman, H. (1994). *İslam açısından borsa*. İstanbul: Ensar Neşriyat.
- Kılıçarslan, S., Adem, K. ve Cömert, O. (2019). Parçacık sürü optimizasyonu kullanılarak boyutu azaltılmış mikrodizi verileri üzerinde makine öğrenmesi yöntemleri ile prostat kanseri teşhisi. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 7, 769- 777.
- Kırelli, Y. (2016). *E-Ticaret siteleri için sahtekarlık tespit sistemleri*. (Yüksek Lisans Tezi), İstanbul ticaret üniversitesi, Bilgisayar Mühendisliği Ana Bilim Dalı, İstanbul.
- Kızılkaya, Y.M., Oğuzlar, A. (2018). Bazı denetimli öğrenme algoritmalarının R programlama dili ile kıyaslanması. *Karadeniz*, 2018, 37.
- Kızrak, A. (2018). “Yapay zeka ve derin öğrenmeye başlama rehberi”, <https://medium.com/@ayyucekizrak/yapay-zekaya> ba% C5% 9Flamarehberi91e79d3de8e1, (Erişim tarihi: 29 Eylül 2021).
- Koçoğlu, F.Ö. (2012). *Veri madenciliği sürecinde veri ayırıklaştırma yöntemlerinin karşılaştırılması ve bir uygulama*. (Yüksek Lisans Tezi), İstanbul Üniversitesi Enformatik Ana Bilim Dalı, İstanbul.
- Köktürk, F., Ankaralı, H., Sümbüloğlu, V. (2009). “Veri madenciliği yöntemlerine genel bakış”. *Türkiye Klinikleri Journal of Biostatistics*, 1, 20-25.

- Krauss, C., Do, X. A., ve Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forest: statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689-702.
- Leech, N. L., Barrett, K. C. ve Morgan, G. A. (2004). *Spss for intermediate statistics: use and interpretation*. Manwah New Jersey: Lawrence Erlbaum Associates Publishers.
- Mete, S. (2019). *Emosyonel uyarılara karşı göz hareketlerinin sınıflandırılması*. (Yüksek Lisans Tezi), Altınbaş Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Nahil, A., ve Lyhyaoui, A. (2018). Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of casablanca stock Exchange. *Procedia Computer Science*, 127, 161-169.
- Ordu, Akkaya, B. M. (2020). Emtia finansallaşmasında borsa İstanbul ve petrol bağıntısı açısından bir bakış. *Ankara Üniversitesi SBF Dergisi*, 75(2), 687-714.
- Öztürk, K. ve Şahin, M. E. (2018). Yapay sinir ağları ve yapay zekaya genel bir bakış. *Takvim-i Vekayi*, 6(2), s. 25-36.
- Pabuçcu, H. (2019). Borsa endeks hareketlerinin makine öğrenmesi algoritmaları ile tahmini. *UIİD, IJEAS*, 23, 179-190.
- Patel, J., Shah, S., Thakkar, P. ve Kotecha, K. (2015). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications*, 42, 259 -268.
- Sağbaş, E. A. ve Ballı, S. (2016). *Akıllı telefon sensör verileri ile eylem tanımda lojistik regresyon ve knn yöntemlerinin karşılaştırılması*. 1st International Conference on Engineering Technology and Applied Science, 22, 894-898.
- Sakarya, Ş. ve Yılmaz, Ü. (2019). Derin öğrenme mimarisi kullanılarak BİST30 indeksinin tahmini. *European Journal of Educational & Social Sciences*, 4(2), 2654-6621.
- Saraç, T. (2004). *Yapay sinir ağları*. Basılmamış Seminer Projesi, Gazi Üniversitesi Endüstri Mühendisliği Bölümü, Ankara.

- Sarıkaya, B. (2021). *İstatistiksel analiz yöntemleri ve makine öğrenme yöntemleri ile film başarı tahmini*. (Tezli Yüksek Lisans), Başkent Üniversitesi, Bilgisayar Mühendisliği Ana Bilim Dalı, Ankara.
- Sayın, A., Yandı, A., Oyar, E. (2017). Kayıp veri ile baş etme yöntemlerinin madde parametrelerine etkisinin incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(4), 490-510.
- Sezer, Ö. B. (2018). *Zaman serisi verilerinin derin yapay sinir ağları ile analizi ve en iyilenmesi: Finansal tahmin algoritmaları*. (Doktora Tezi), TOBB Ekonomi ve Teknoloji Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Shen, F., Chao, J., ve Zhao, J. (2015). *Forecasting exchange rate using deep belief networks and conjugate gradient method*. In: *Neurocomputing* 167, 243-253.
- Soy Temür, A., (2019). *İşletmelerin satış bütçelerinin oluşturulmasında ARIMA, LSTM, Hibrit Modellerin karşılaştırılması: Üretim işletmesi Örneği*. (Doktora Tezi), Sakarya Üniversitesi, Sosyal Bilimler Enstitüsü, Sakarya.
- Staudemeyer, R. C. ve Rothstein Morris, E. (2019). *Understanding LSTM a tutorial into long-short term memory recurrent neural networks*. Faculty of Computer Science Schmalkalden University of Applied Sciences, Germany.
- Şişmanoğlu, G., Koçer, F., Önde, M.A., Şahinsöz, Ö. K. (2020). Derin öğrenme yöntemleri ile borsada fiyat tahmini. *BEU Fen Bilimleri Dergisi*, 9(1), 434-455
- Turan, S. (2019). *Uzun Kısa Süreli Hafıza ve Geçitli Yinelenen Birim İle Borsa İstanbul 100 Endeks Değeri Tahmini Üzerine Bir Uygulama*. (Yüksek Lisans Tezi), Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Uğur, A. (2009). *Hisse Senedi Getirilerinin Panel Veri Analizi ile Tahmini: İstanbul Menkul Kıymetler Borsasında Bir İnceleme*. (Doktora Tezi), İnönü Üniversitesi, Sosyal Bilimler Enstitüsü, Malatya.
- Vursavuş, K. ve Üçkardeş, F. (2015). K-star, rastgele orman ve karar ağacı (C.4.5) sınıflandırma algoritmaları ile domatesin renk olgunluğu üzerinde bazı mekanik özelliklerin etkisinin belirlenmesi, *Article in Turkish Journal of Agriculture Food Science and Technology*.

9. TEZ DEĞERLENDİRME FORMU

Aksaray Üniversitesi Tez Değerlendirme Formu

Öğrencinin Adı Soyadı: Serap AKBULUT		EVET
Kapak		
1	Tez Başlığı tutanaktaki başlıkla aynı mı?	EVET
2	Kapakdaki ay ve yıl savunmaya girilen tarihle tutarlı mı?	EVET
3	Kapak format kılavuzdaki kapak formatına uygun mu?	EVET
4	Kapakta yazılan tüm yazılar doğru olarak verilmiş mi?	EVET
İçindekiler		
5	Sayfa numaraları tam verilmiş mi?	EVET
6	Şekil, Çizelge vb. listeleri verilmiş mi? Sıralaması doğru mu?	EVET
7	Özet, Abstract, Giriş, Sonuçlar vb. bölümler var mı?	EVET
8	Yazım hataları kontrolü yapıldı mı?	EVET
Giriş		
9	Hazırlanan tezin önemini anlatıyor mu?	EVET
10	İkinci ve Üçüncü dereceden başlık içermemeli kuralına uyuldu mu?	EVET
Özet/Abstract		
11	Kılavuza uygun mu?	EVET
12	Ay ve yıl savunmaya girilen tarihle tutarlı mı?	EVET
13	Özet; tek sayfa, tek aralık, tek paragraf kuralına uygun olarak yazıldı mı?	EVET
14	Bilim kodu, sayfa adedi, anahtar kelimeler ve tez danışmanı yazıldı mı?	EVET
Kaynakça		
15	Kaynakların tamamına metin içinde atıf yapıldı mı?	EVET
16	Kaynak formatı Kılavuzdaki kaynak formatına uygun olarak hazırlanmış mı?	EVET
17	Atıf formatı kılavuzdaki atıf formatına uygun mu?	EVET
Atıf Yöntemi APA 6 CMS ☐ İSNAD ☐		
Genel Değerlendirme		
18	Etik Beyan açıklaması okundu, uyuldu ve imzalandı mı?	EVET
19	Kabul/Onay sayfası kılavuzdaki formata uygun olarak düzenlenmiş mi?	EVET
20	Kabul /Onay sayfasında belirtilen oy birliği/oy çokluğu seçeneklerinden uygun olanı savunmayla tutarlı olacak şekilde belirlenmiş mi?	EVET
21	Sayfa kenar boşluklar ve sayfa numaraları kılavuzdaki formatına uygun mu?	EVET
22	Paragraf boşlukları ve metin satır aralığı kılavuza uygun olacak şekilde düzenlenmiş mi?	EVET
23	Başlık yazımları kılavuzdaki başlık formatlarına uygun mu?	EVET
24	Yazı tipi ve boyutu kılavuzdaki yazı tipi ve boyutu formatına uygun mu?	EVET
25	Şekil, Çizelge vb. açıklama ve numaralandırmaları kılavuzdaki formata uygun olarak yazılmış	EVET

Bu tezin tarafımdan “Tez yazım kuralları” okunarak dikkatlice hazırlanmış olduğunu ve doğabilecek her türlü olumsuzluktan sorumlu olacağımı kabul ederim.

Serap AKBULUT